

**TRANSIENT LINEAR PRICE IMPACT**  
**A second generation of market impact models**

by

**KLAAS DE JONG**

***THESIS***  
*for the degree of*  
***MASTER OF SCIENCE***

*(Master in Modeling and Data Analysis - MOD5930)*



*Faculty of Mathematics and Natural Sciences*  
*University of Oslo*

*May 2015*

© Klaas de Jong, 2015

Transient linear price impact. A second generation of market impact models.

Klaas de Jong

<http://www.duo.uio.no/>

Printed by Reprosentralen, University of Oslo

## Abstract

In classic mathematical finance, a trader's actions have no direct influence on the asset price. For small trades this is a reasonable assumption, but large trades fire back at the underlying price. We consider a transient linear price impact model in discrete time, and find a deterministic and unique optimal trading strategy when the decay of price impact is given as a positive-definite quadratic form. Examples of the associated so-called resilience functions show a new type of price manipulation, which will be called transaction-triggered price manipulation. To exclude this kind of price manipulation, convexity of the resilience function appears to be both necessary and sufficient. Since nonconstant, convex functions generate positive definite quadratic forms, standard price manipulation is excluded in this case as well. The effects of risk aversion can be handled similarly to the way the standard optimal order execution problem is solved. The discrete-time model can be extended to continuous time, and we find some similar results. It appears that optimal strategies can be characterized as measure-valued solutions of a generalized Fredholm integral equation of the first kind. However, to guarantee the existence of an optimal trading strategy, positive definiteness does not hold, and we need convexity of the decay kernel. As in the discrete-time case, this excludes the existence of transaction-triggered price manipulation strategies.

**Keywords:** market impact model, transient price impact, optimal order execution, positive (semi-)definite quadratic form, price manipulation, Lagrange multiplier, Bochner's theorem, Fourier transform, convex function, Lebesgue-Stieltjes measure, Lebesgue-Stieltjes integral, integration by parts for Lebesgue-Stieltjes integrals, dominated convergence, continuity theorem, portmanteau theorem, Fubini-Tonelli theorem, transaction-triggered price manipulation, risk aversion, bilinear form, polarization, Fredholm integral equation, Borel probability measure, compact metric space, Prohorov's theorem.





# Contents

<b>Preface</b>	<b>v</b>
<b>1 Introduction to market impact modeling</b>	<b>1</b>
<b>2 A transient linear model in discrete time</b>	<b>3</b>
2.1 Introducing a second-generation model . . . . .	3
2.2 Expected execution costs . . . . .	5
2.3 An optimal trading strategy . . . . .	7
<b>3 Examples of resilience functions</b>	<b>11</b>
3.1 Positive-definite resilience functions . . . . .	11
3.2 Convex resilience functions . . . . .	15
3.3 Non-convex resilience functions . . . . .	26
<b>4 Transaction-triggered price manipulation</b>	<b>31</b>
4.1 Sufficiency of convexity . . . . .	32
4.2 Necessity of convexity . . . . .	37
<b>5 Some properties of optimal strategies</b>	<b>41</b>
<b>6 Risk aversion</b>	<b>45</b>
6.1 A mean-variance functional . . . . .	45
6.2 Excluding transaction-triggered price manipulation . . . . .	48
<b>7 A transient linear model in continuous time</b>	<b>51</b>
7.1 Replicating the discrete second-generation model . . . . .	51
7.2 Expected execution costs . . . . .	53
7.3 A Fredholm integral equation . . . . .	55
7.4 Nonexistence of optimal solutions . . . . .	64
<b>8 A well-behaved optimal trading strategy</b>	<b>69</b>

<b>Conclusion</b>	<b>73</b>
<b>Recommendations</b>	<b>75</b>
<b>A R code</b>	<b>77</b>
<b>Bibliography</b>	<b>83</b>

# Preface

*“From an economic perspective, market impact is one of the basic mechanisms responsible for price formation, and so its analysis might contribute to new insights on how financial markets function.”* - A. Schied & A. Slynko (2011)

The topic of market impact risk (and more generally, liquidity risk) has grown to gain a lot of attention, and many works have been published in recent years. For this thesis, the survey paper by Schied and Slynko (2011, [30]) is used as a starting point. We provide a systematic presentation of the theory of second-generation market impact models with transient linear price impact, including complete proofs.

The structure of this thesis is as follows: In Chapter 1 we take a look at the concept of market impact. Then, in Chapter 2, we consider a discrete-time model and deduce an optimal trading strategy, which will be given explicitly. In Chapter 3 we study some examples of resilience functions. These examples show a new type of market irregularities, which we deal with in Chapter 4. Next, in Chapter 5, we discuss some properties of optimal strategies, after which we study the concept of risk aversion in Chapter 6. Then, in Chapter 7, we consider a model in continuous time, and find necessary conditions for the existence of an optimal order execution strategy. Finally, we state sufficient conditions for the existence of such an optimal strategy in Chapter 8. These conditions appear to be sufficient for the optimal strategy to be well-behaved.

The first part of this thesis, about models in discrete time (Chapters 2 to 6), is mainly based on the papers by Alfonsi et al. (2012, [4]) and Schied & Slynko (2011, [30]). The propositions and theorems are taken from [4], and the proofs are based on sketches presented in that paper. These sketches are elaborated, extended with complete arguments and presented in a way that they are understandable for a reader without much experience in this field. The examples in Chapter 3 and 6 are taken from both [4] and [30], but they are extended with complete computations. All figures used to support the examples are generated by self-written code (cf. Appendix A).

For the theory on continuous time models (Chapters 7 and 8), we have mainly used the paper by Gatheral et al. (2012, [18]). All propositions, theorems and proofs are taken from this paper, the proofs being adapted and extended with complete arguments. Since the main purpose of this thesis was to cover models in discrete-time, the proofs of some of the propositions in the continuous-time part are skipped or only a sketch is provided. The examples in this part are also taken from [18], and extended with complete computations and explanatory figures, which are generated by self-written code (cf. Appendix A).

Perhaps surprisingly, much of the theory will be about deterministic quantities, because (as we will see) stochastic terms cancel. This leads to elegant solutions for many of the problems. As a consequence, basic knowledge will be sufficient to understand most of the theory, especially when it comes to discrete-time modeling.

Finally, I want to thank everyone who, in any way, helped me to write this thesis. Special thanks go to my supervisor, professor Tom Lindstrøm, for all his help and guidance, by answering my many questions and giving advice on how to continue when I was stuck. I am also grateful to my fiancée, Katrine Anthonisen, for her never ending support, patience and determination to motivate me.

*Klaas de Jong*  
Oslo, May 2015

# Chapter 1

## Introduction to market impact modeling

Standard asset price models assume that a trader's actions have no direct influence on the asset price. This means that a trader can buy and sell unlimited amounts of assets for the current market price. When dealing with smaller traders, this is a reasonable assumption. However, there is usually only a small group of buyers who are interested in paying the current price, and in order to carry out a large selling order, one has to lower one's price.

According to Schied & Slynko (2011, [30]), market impact risk is a specific kind of liquidity risk, describing “the risk of not being able to execute a trade at the currently quoted price because this trade feeds back in an unfavorable manner on the underlying price”. As an example of a case in which market impact risk played an important role, they mention the LTCM crisis in 1998, when the Russian financial crisis triggered panicked investors to sell their bonds, leading to a significant price fall. But market impact risk plays a role in much smaller trades as well, and “it belongs to the daily business of many financial institutions” ([30]). This happens through implementation of market impact risk in today's automated trading algorithms, using market impact models, which take into account how trading strategies temporarily effect asset prices.

The core idea behind these market impact models is the observation that liquidity costs can be reduced by splitting a large trade into smaller trades, called child orders, to be executed over a given time interval. In this way, after each child order is carried out, the market has time to (partly) recover before the next child order is executed.

The first generation of market impact models distinguished between a temporary and a permanent price impact component, the former only affecting the trade that had triggered it, while the latter component affected all current and future trades equally. Writing  $S^0$  for the unaffected stock price process (the process driven by the actions of noise traders, usually assumed to be a martingale), these first-generation market impact models are of the form

$$S_t = S_t^0 + \eta \dot{X}_t + \gamma (X_t - X_0),$$

where  $X_t$  denotes the number of shares in the trader's portfolio, and  $\eta$  and  $\gamma$  are constants ([30]). The term  $\eta \dot{X}_t$  corresponds to trading  $\dot{X}_t dt$  shares at time  $t$ . This temporary impact only plays a role while a trade is executed, and hence only affects the current order. The term  $\gamma (X_t - X_0)$  describes the effect of all accumulated transactions up to time  $t$ , which is permanent.

According to [30], the distinction between temporary and permanent price impact is only reasonable as long as the time between individual child orders is long enough. Empirical studies show that on a finer time scale, the execution of a child order causes an immediate price impact which decays over time<sup>1</sup>, meaning that price impact is (at least partly) temporary. Because of the high trading frequency used in today's electronic trading systems, the distinction between a temporary and a permanent price impact component can no longer be considered realistic. To deal with this, the decay of price impact is modeled explicitly in a second generation of market impact models. These are the models we will consider in this paper.

---

<sup>1</sup>Note that this price impact decay is the very reason of the fact that splitting up an order into smaller child orders can reduce costs.

## Chapter 2

# A transient linear model in discrete time

In this chapter, we take a look at the second-generation market impact model introduced in [4]. It describes asset prices for a large trader who can move those prices. We assume that, as long as this trader is inactive, the asset prices are described by a martingale  $(S_t^0)_{t \geq 0}$ , defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . The major reason for requiring  $S^0$  to be a martingale, is the fact that a nonzero drift would cause arbitrage opportunities. Another reason is that, due to the typically short trading horizons, drift effects are usually ignored in the market impact literature ([4]).

### 2.1 Introducing a second-generation model

To buy or liquidate a portfolio of a given number of shares  $X_0$ , the trader can use a *strategy*  $\xi = (\xi_{t_0}, \dots, \xi_{t_N})$ , consisting of trades  $\xi_{t_n}$  at predetermined trading times  $t_n$ , where  $0 \leq t_0 < t_1 < \dots < t_N = T$ . We thus assume that

$$\xi^T \mathbf{1} = \sum_{n=0}^N \xi_{t_n} = X_0.$$

When  $X_0 > 0$ , a buy program is executed, whereas  $X_0 < 0$  corresponds to a sell program. A strategy satisfying  $X_0 = 0$  is called a *round trip*. We assume that  $\xi$  is adapted, so each trade  $\xi_{t_n}$  is allowed to depend on all information available to the trader at time  $t_n$ . We also assume that  $\xi$  is bounded. From an economic perspective, this can be done without loss of generality, since there exists only a limited number of shares of each asset.

We are now ready to introduce the model: under the strategy  $\xi = (\xi_{t_0}, \dots, \xi_{t_N})$ , the asset price at time  $t$  is given by

$$S_t = S_t^0 + \sum_{t_n < t} \xi_{t_n} G(t - t_n), \quad (2.1)$$

where  $G : [0, \infty) \rightarrow [0, \infty)$  is a nonincreasing *resilience function*, describing the price impact of a trade at  $t_n$  over time. From this model, we see that price impact is assumed to be linear. The two main reasons for this are mathematical tractability and the fact that it is still an open problem how to best model nonlinear transient impact ([4]). So far, there is no mathematical need to require  $G$  to be nonincreasing. However, intuitively,

this makes most sense, and we will see that we need the assumption in most of the theory that will follow. Since  $G$  is nonnegative, we see that buying  $\xi_{t_n}$  shares at  $t_n$  leads to a price increase, while selling  $\xi_{t_n}$  shares at  $t_n$  causes a price decrease.

To illustrate the impact of a trade at  $t_n$ , let us have a look at the price change in the case in which there is only one trade (at  $t_n$ ):

$$S_t - S_t^0 = \begin{cases} 0 & t \leq t_n, \\ \xi_{t_n} G(t - t_n) & t > t_n. \end{cases}$$

Its graph may look like the left-hand side of Figure 2.1. The immediate price impact of the order  $\xi_{t_n}$  for the trader is  $\xi_{t_n} G(0)$ . That is, the price moves from  $S_{t_n}$  to  $S_{t_n+} := S_{t_n} + \xi_{t_n} G(0)$  when the trade  $\xi_{t_n}$  is executed at  $t_n$ . From (2.1), we see that  $G(0)$  does not enter the model. This implies that it is only of importance for the current trade, and does not influence later trades. The difference between  $G(0)$  and  $G(0+) := \lim_{t \downarrow 0} G(t)$  for functions  $G$  which are not continuous at time 0, can be caused by for instance transaction costs. As shown in Figure 2.1, the market impact of a trade  $\xi_{t_n}$  at  $t_n$  is built up by the following three types of price impact:

- the *instantaneous impact*  $\xi_{t_n}(G(0) - G(0+))$ , describing what the trader pays extra (for example in the form of transaction costs) compared to the market price right after the trade is executed,
- the *permanent impact*  $\xi_{t_n} G(\infty)$ , where  $G(\infty) := \lim_{t \rightarrow \infty} G(t)$ , describing the price impact in the long run, and affecting all future and current trades equally,
- the *transient impact*  $\xi_{t_n}(G(0+) - G(\infty))$ , describing the part of the price impact that decays over time.

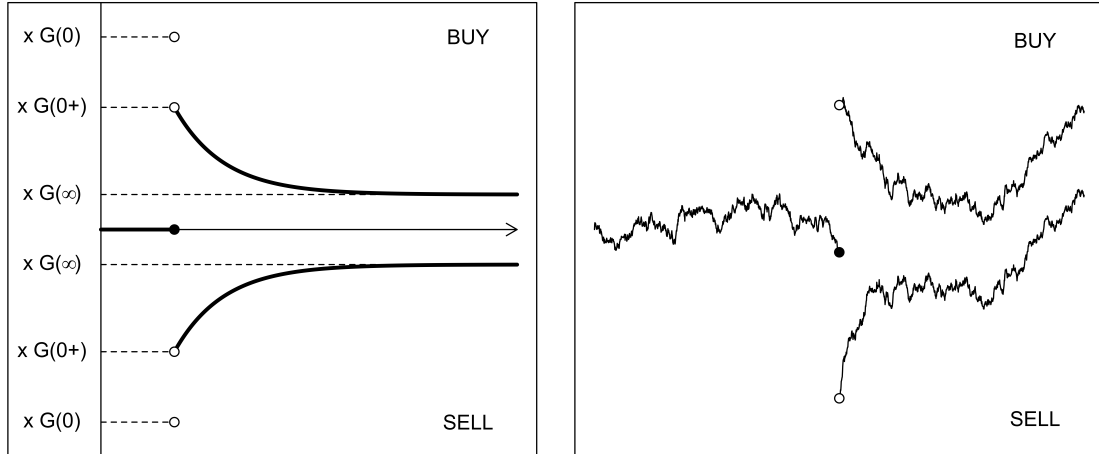


Figure 2.1: Price impact of a single trade. To the left a graph of  $S_t - S_t^0$ , to the right a graph of  $S_t$ .



## 2.2 Expected execution costs

We now want to define the expected execution cost of a strategy  $\xi$ . Every trade  $\xi_{t_n}$  causes a price movement from  $S_{t_n}$  to  $S_{t_n+} = S_{t_n} + \xi_{t_n}G(0)$ . In order to compute the execution cost of the trade  $\xi_{t_n}$ , we can think of it as existing of many infinitesimal investments, which we assume to be of the same size  $d\xi$ . For these infinitesimal investments, the asset price change is given by

$$dS = G(0) d\xi,$$

from which immediately follows that

$$d\xi = \frac{1}{G(0)} dS.$$

So at each price  $S$ ,  $G(0)^{-1}dS$  shares are available for buying or selling. Integrating over all infinitesimal investments, or equivalently, over all available prices, we find that

$$\int_0^{\xi_{t_n}} S d\xi = \int_{S_{t_n}}^{S_{t_n+}} \frac{S}{G(0)} dS = \frac{1}{2G(0)} (S_{t_n+}^2 - S_{t_n}^2) = \frac{1}{G(0)} \frac{S_{t_n} + S_{t_n+}}{2} (S_{t_n+} - S_{t_n}), \quad (2.2)$$

where we recognize the average price  $\frac{1}{2}(S_{t_n} + S_{t_n+})$  in the last expression. Since the execution of a so-called block order typically only takes some days or even hours, we can neglect the possibility that negative prices occur, even when the model allows for such prices. Therefore, the quantity expressed in (2.2) is positive for buy orders  $\xi_{t_n} > 0$  and negative for sell orders  $\xi_{t_n} < 0$ . It can thus be regarded as the cost of the trade  $\xi_{t_n}$ : the amount of money one pays for buying  $\xi_{t_n}$  shares, or in the case of selling, the amount of money one “pays” for selling  $\xi_{t_n}$  shares (that is, its absolute value is what one receives for selling). Using this, we can define the expected execution cost  $\mathcal{C}(\xi)$  of the strategy  $\xi$  as the expected cumulative costs of all trades in the strategy  $\xi$ :

$$\mathcal{C}(\xi) := \mathbb{E} \left[ \sum_{n=0}^N \int_{S_{t_n}}^{S_{t_n+}} \frac{S}{G(0)} dS \right] = \frac{1}{2G(0)} \mathbb{E} \left[ \sum_{n=0}^N (S_{t_n+}^2 - S_{t_n}^2) \right]. \quad (2.3)$$

The *optimal order execution problem* now corresponds to minimizing  $\mathcal{C}(\xi)$  for a given  $X_0$ . Defining  $M$  as the matrix with entries  $M_{ij} = G(|t_i - t_j|)$ , we have the following important result:

**Proposition 1.** ([4]) *The expected execution cost of a strategy  $\xi$  is given by*

$$\mathcal{C}(\xi) = X_0 S_0 + \mathbb{E} [C(\xi)], \quad (2.4)$$

where  $C$  is the quadratic form

$$C(\mathbf{x}) := \frac{1}{2} \mathbf{x}^T M \mathbf{x} = \frac{1}{2} \sum_{i,j=0}^N x_i x_j G(|t_i - t_j|),$$

for  $\mathbf{x} = (x_0, \dots, x_N) \in \mathbb{R}^{N+1}$ .

*Proof.* First note that

$$\begin{aligned}
\mathbf{x}^T M \mathbf{x} &= (x_0, \dots, x_N) \begin{pmatrix} G(|t_0 - t_0|) & \cdots & G(|t_0 - t_N|) \\ \vdots & \ddots & \vdots \\ G(|t_N - t_0|) & \cdots & G(|t_N - t_N|) \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ x_N \end{pmatrix} \\
&= (x_0, \dots, x_N) \begin{pmatrix} \sum_{j=0}^N x_j G(|t_0 - t_j|) \\ \vdots \\ \sum_{j=0}^N x_j G(|t_N - t_j|) \end{pmatrix} \\
&= \sum_{i=0}^N x_i \left( \sum_{j=0}^N x_j G(|t_i - t_j|) \right) = \sum_{i,j=0}^N x_i x_j G(|t_i - t_j|),
\end{aligned}$$

which proves the equality between the two expressions for  $C(\mathbf{x})$ . Next, consider  $\mathcal{C}(\boldsymbol{\xi})$ . Since, by the definition of  $S_{t_n+}$ ,

$$S_{t_n+}^2 - S_{t_n}^2 = (S_{t_n} + \xi_{t_n} G(0))^2 - S_{t_n}^2 = 2 S_{t_n} \xi_{t_n} G(0) + \xi_{t_n}^2 G(0)^2,$$

it follows that

$$\begin{aligned}
\frac{1}{2G(0)} \sum_{n=0}^N (S_{t_n+}^2 - S_{t_n}^2) &= \sum_{n=0}^N \left( S_{t_n} \xi_{t_n} + \frac{G(0)}{2} \xi_{t_n}^2 \right) \\
&= \sum_{n=0}^N \left( \frac{G(0)}{2} \xi_{t_n}^2 + \xi_{t_n} \left( S_{t_n}^0 + \sum_{t_k < t_n} \xi_{t_k} G(t_n - t_k) \right) \right) \\
&= \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 + \frac{1}{2} \sum_{n=0}^N \left( \xi_{t_n}^2 G(0) + 2 \sum_{t_k < t_n} \xi_{t_n} \xi_{t_k} G(|t_n - t_k|) \right).
\end{aligned}$$

Note that  $\xi_{t_n} \xi_{t_k} G(|t_n - t_k|)$  is symmetric in  $t_k$  and  $t_n$ . This means that we can rewrite the expression as

$$\begin{aligned}
\frac{1}{2G(0)} \sum_{n=0}^N (S_{t_n+}^2 - S_{t_n}^2) &= \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \\
&+ \frac{1}{2} \sum_{n=0}^N \left( \sum_{t_k = t_n} \xi_{t_n} \xi_{t_k} G(|t_n - t_k|) + \sum_{t_k < t_n} \xi_{t_n} \xi_{t_k} G(|t_n - t_k|) + \sum_{t_k > t_n} \xi_{t_n} \xi_{t_k} G(|t_n - t_k|) \right) \\
&= \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 + \frac{1}{2} \sum_{k,n=0}^N \xi_{t_n} \xi_{t_k} G(|t_n - t_k|) \\
&= \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 + C(\boldsymbol{\xi}).
\end{aligned}$$

Now since  $(S_t^0)_{t \geq 0}$  is a martingale, its expected value at time  $T$  is finite, and hence by the law of total expectation,

$$\mathbb{E} \left[ \sum_{n=0}^N \xi_{t_n} S_T^0 \right] = \sum_{n=0}^N \mathbb{E} \left[ \mathbb{E} \left[ \xi_{t_n} S_T^0 \mid \mathcal{F}_{t_n} \right] \right],$$

where  $(\mathcal{F}_t)_{t \geq 0}$  is the filtration associated with the probability space on which  $(S_t^0)_{t \geq 0}$  was defined. Using that  $\xi$  is adapted and  $(S_t^0)_{t \geq 0}$  is a martingale, we see that this equals

$$\sum_{n=0}^N \mathbb{E} \left[ \xi_{t_n} \mathbb{E} \left[ S_T^0 \mid \mathcal{F}_{t_n} \right] \right] = \mathbb{E} \left[ \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \right].$$

Hence

$$\mathbb{E} \left[ \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \right] = \left( \sum_{n=0}^N \xi_{t_n} \right) \mathbb{E} \left[ S_T^0 \right] = X_0 S_0^0, \quad (2.5)$$

as  $\sum_{n=0}^N \xi_{t_n} = X_0$  is known. Clearly,  $S_0^0 = S_0$ , which gives us the final result:

$$\mathcal{C}(\xi) = \mathbb{E} \left[ \frac{1}{2G(0)} \sum_{n=0}^N (S_{t_{n+}}^2 - S_{t_n}^2) \right] = X_0 S_0 + \mathbb{E} [C(\xi)]. \quad \square$$

## 2.3 An optimal trading strategy

The significance of Proposition 1 lies in the fact that all stochastic  $S_t$ -terms have disappeared, essentially because  $(S_t^0)_{t \geq 0}$  is a martingale and the price impact is linear. Therefore, the optimal strategy  $\xi^*$  does not depend on the price process, and becomes deterministic. It follows that if  $\mathbf{x}^* \in \mathbb{R}^{N+1}$  minimizes  $C(\mathbf{x})$  in  $\{\mathbf{x} \in \mathbb{R}^{N+1} \mid \mathbf{x}^T \mathbf{1} = X_0\}$ , then the deterministic strategy  $\xi^* := \mathbf{x}^*$  minimizes the expected execution costs over all available strategies. Since  $C$  is a quadratic form, such a minimizer  $\mathbf{x}^*$  clearly exists as soon as  $C(\mathbf{x}) \geq 0$  for all  $\mathbf{x}$ . Recall that  $G$  (or equivalently,  $M$ ) is called positive semidefinite if  $C(\cdot) \geq 0$  for all possible time grids  $0 = t_0 \leq t_1 \leq \dots \leq t_N = T$  and  $N \in \mathbb{N}$ . When even  $C(\mathbf{x}) > 0$  for  $\mathbf{x} \neq \mathbf{0}$ ,  $G$  (or equivalently,  $M$ ) is called (strictly) positive definite. This means that when  $G$  is positive definite,  $\xi = \mathbf{0}$  is the unique optimal strategy in the class of round trips. We can also say that for positive-definite  $G$ , price manipulation is excluded.

So what is price manipulation exactly? For standard asset pricing models, the absence of arbitrage is guaranteed by the existence of an equivalent martingale measure. For market impact models however, requiring the unaffected stock price process  $S^0$  to be a martingale, is not always enough. There may still exist strategies which, when suitably rescaled and repeated, can lead to a weak form of arbitrage ([4]). These *price manipulation strategies* are round trips  $\xi$  with strictly negative expected execution costs  $\mathcal{C}(\xi) < 0$ . In other words, they are strategies for which money is earned on average, while the trader ends up with the same portfolio as he started with. We quote from [4]: “In a market impact model that admits price manipulation, efficient martingale dynamics can be turned into stock price behavior that is favorable to a large trader. For such traders, or for their automated trading programs, using a model that admits price manipulation strategies thus provides an incentive to play risky strategies so as to make profit on average.” For this reason, we want to exclude price manipulation from our model.

The following theorem gives conditions under which the existence of price manipulation strategies can be ruled out. Moreover, it describes a unique optimal strategy, also for the case  $X_0 \neq 0$ .

**Theorem 1.** ([4]) *If  $G$  is positive semidefinite, the model does not admit price manipulation strategies. Moreover, optimal order execution strategies exist but need not be unique.*

*If  $G$  is positive definite, then for every initial position  $X_0$  there exists a unique optimal order execution strategy  $\xi^*$  given by*

$$\xi^* = \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1}. \quad (2.6)$$

In [4] it says that this theorem follows immediately from  $G$  being positive (semi-)definite, but this might not be clear for everyone. We will therefore give a proof here.

*Proof.* If  $G$  is positive semidefinite,  $C(\mathbf{x}) \geq 0$  for all  $\mathbf{x}$ . We already noticed that optimal order execution strategies exist in this situation. When dealing with a round trip ( $X_0 = 0$ ),  $C(\mathbf{x}) \geq 0$  if and only if  $\mathcal{C}(\mathbf{x}) \geq 0$ , which follows immediately from (2.4). Hence the model does not allow for price manipulation (recall that a price manipulation strategy was defined as a round trip with strictly negative expected execution cost).

If  $G$  is positive definite, we can prove that  $M$  is invertible. For a positive definite matrix  $M$ ,  $\mathbf{x}^t M \mathbf{x} > 0$  for all  $\mathbf{x} \neq \mathbf{0}$ . In particular, if  $\mathbf{v} \neq \mathbf{0}$  is an eigenvector of  $M$ , then  $\mathbf{v}^t M \mathbf{v} = \mathbf{v}^t \mathbf{v} \lambda = \|\mathbf{v}\|^2 \lambda > 0$ , where  $\|\cdot\|$  denotes the Euclidean norm. It follows that  $\lambda > 0$ , which means that all eigenvalues of  $M$  are positive. In particular, zero is not an eigenvalue of  $M$ , so the equation  $M\mathbf{x} = \mathbf{0} = \mathbf{0} \cdot \mathbf{x}$  only has the trivial solution  $\mathbf{x} = \mathbf{0}$ . This means that  $M$  is invertible.

We already know that a minimizer of  $C$  exists, but we do not know whether it is unique. Let us take an arbitrary minimizer of  $C$  whose components sum up to  $X_0$ , and call it  $\mathbf{x}^0$ . Then we define

$$f : \mathbb{R}^{N+1} \rightarrow \mathbb{R}, \quad f(\mathbf{x}) := C(\mathbf{x}) = \frac{1}{2} \sum_{i,j=0}^N x_i x_j M_{ij},$$

$$g : \mathbb{R}^{N+1} \rightarrow \mathbb{R}, \quad g(\mathbf{x}) := \mathbf{1}^T \mathbf{x} = \sum_{j=0}^N x_j,$$

where  $M_{ij}$  is the  $i, j$ th element of the matrix  $M$ . Let  $S = \{\mathbf{x} \in \mathbb{R}^{N+1} \mid g(\mathbf{x}) = X_0\}$ , and note that  $\mathbf{x}^0 \in S$ . Now since  $f|_S$  (“ $f$  restricted to  $S$ ”) has a local minimum on  $S$  at  $\mathbf{x}^0$ , and  $\nabla g(\mathbf{x}^0) = \mathbf{1} \neq \mathbf{0}$ , we know by the method of Lagrange multipliers (cf. e.g. [22]) that there exists  $\lambda \in \mathbb{R}$  such that

$$\nabla f(\mathbf{x}^0) = \lambda \cdot \nabla g(\mathbf{x}^0) = \lambda \cdot \mathbf{1}. \quad (2.7)$$

Writing

$$f(\mathbf{x}) = \frac{1}{2} \left( x_0 \sum_{j=0}^N x_j M_{0j} + \dots + x_N \sum_{j=0}^N x_j M_{Nj} \right),$$

we see that

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \frac{1}{2} \left( \sum_{\substack{j=0 \\ j \neq i}}^N x_j M_{ij} + \sum_{\substack{j=0 \\ j \neq i}}^N x_j M_{ji} + 2 x_i M_{ii} \right) = \sum_{j=0}^N x_j M_{ij},$$

and hence

$$\nabla f(\mathbf{x}) = \left( \sum_{j=0}^N x_j M_{0j}, \dots, \sum_{j=0}^N x_j M_{Nj} \right) = M\mathbf{x}.$$

It thus follows from (2.7) that

$$\mathbf{x}^0 = \lambda \cdot M^{-1}\mathbf{1},$$

a unique solution. But this implies that

$$g(\mathbf{x}^0) = g(\lambda \cdot M^{-1}\mathbf{1}) = \lambda \cdot \mathbf{1}^T M^{-1}\mathbf{1} = X_0.$$

As we saw,  $M$  has only positive eigenvalues. Since for invertible  $M$ ,  $M\mathbf{v} = \lambda\mathbf{v}$  if and only if  $\mathbf{v} = M^{-1}M\mathbf{v} = \lambda M^{-1}\mathbf{v}$ , this implies that  $M^{-1}$  has only positive eigenvalues as well. Now  $M^{-1}$  is symmetric and hence diagonalizable, which means that  $\mathbb{R}^{N+1}$  has a basis consisting of eigenvectors  $v_0, \dots, v_N$  of  $M^{-1}$ . We can thus decompose any vector  $\mathbf{x} \in \mathbb{R}^{N+1}$  as  $\mathbf{x} = \sum_i \alpha_i \mathbf{v}_i$ . It follows that for all  $\mathbf{x} \neq \mathbf{0}$ ,

$$\mathbf{x}^T M^{-1} \mathbf{x} = \left( \sum_{i=0}^N \alpha_i \mathbf{v}_i \right)^T M^{-1} \left( \sum_{j=0}^N \alpha_j \mathbf{v}_j \right) = \sum_{i,j=0}^N (\alpha_i \mathbf{v}_i)^T M^{-1} (\alpha_j \mathbf{v}_j) > 0,$$

since for any  $i \in \{0, \dots, N\}$ ,  $\alpha_i \mathbf{v}_i$  is an eigenvector of  $M^{-1}$ . In particular,  $\mathbf{1}^T M^{-1} \mathbf{1} > 0$ , and it follows that

$$\lambda = \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}}.$$

The unique optimal order execution strategy  $\xi^*$  is therefore given by

$$\xi^* = \mathbf{x}^0 = \lambda \cdot M^{-1} \mathbf{1} = \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1}. \quad \square$$

Due to Theorem 1, we are interested in the class of positive-definite resilience functions. However, it can be quite difficult to determine whether a given function is positive definite, using the definition of positive definiteness. The following theorem by Bochner can be helpful in such situations. It gives a characterization of precisely those resilience functions that are positive (semi-)definite within the class of continuous resilience functions<sup>1</sup>.

**Proposition 2. (Bochner's theorem, [4])** *A continuous resilience function  $G$  is positive semidefinite if and only if the function  $G(|\cdot|)$  is the Fourier transform of a positive finite Borel measure  $\mu$  on  $\mathbb{R}$ , i.e. if and only if*

$$G(|x|) = \int e^{ixy} \mu(dy).$$

*When, in addition, the support of  $\mu$  is not discrete, then  $G$  is even (strictly) positive definite.*

Strictly speaking, the last sentence is not part of Bochner's theorem, but an addition due to [13]. However, for convenience, we will refer to it as a part of Bochner's theorem in the sequel. In [13], by a discrete set in  $\mathbb{R}$  is meant a subset  $D \subset \mathbb{R}$  consisting solely of isolated elements of  $\mathbb{R}$ . That is, for every  $d \in D$  there exists an interval  $(d - r, d + r)$

---

<sup>1</sup>For this thesis, the theory behind Bochner's theorem is not of main importance. The result is merely used as a machinery to point out positive-definite functions. We will therefore skip the proof of Proposition 2.

that does not contain any other elements of  $D$ . The set  $\mathbb{N}^{-1}$  for instance, defined as  $\mathbb{N}^{-1} := \{1/n \mid n \in \mathbb{N} \setminus \{0\}\}$ , is discrete, but its extension  $\mathbb{N}^{-1} \cup \{0\}$  is not. Recall that a Borel measure on  $\mathbb{R}$  is a measure defined on the Borel  $\sigma$ -algebra on  $\mathbb{R}$ , which by its definition is the smallest  $\sigma$ -algebra containing all open sets in  $\mathbb{R}$ .

In the next chapter we will take a look at some examples of resilience functions. They will show that even in the class of positive-definite resilience functions, undesirable behavior may occur. It will appear that requiring the absence of price manipulation is not enough if one wants to model a well-behaved market.

## Chapter 3

# Examples of resilience functions

In this chapter we consider several examples of resilience functions. We start with some examples for which we can use Bochner's theorem to derive optimal strategies.

### 3.1 Positive-definite resilience functions

**Example 1. (Permanent price impact)** The constant resilience function  $G(t) \equiv 1$  is the Fourier transform of the Dirac measure  $\mu := \delta_0$ , given by

$$\delta_0(y) = \begin{cases} 1 & \text{if } y = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, we have for all  $t \in \mathbb{R}$  that

$$\int e^{ity} \delta_0(dy) = e^{it \cdot 0} = 1 = G(|t|).$$

Clearly,  $\mu$  is a finite Borel measure, and it follows from Bochner's theorem that  $G$  is positive semidefinite. For any strategy  $\xi$ , the costs are given by

$$\mathcal{C}(\xi) = X_0 S_0 + \frac{1}{2} \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} G(|t_i - t_j|) = X_0 S_0 + \frac{1}{2} \left( \sum_{i=0}^N \xi_{t_i} \right) \left( \sum_{j=0}^N \xi_{t_j} \right) = X_0 S_0 + \frac{1}{2} X_0^2.$$

Since this is independent of  $\xi$ , all strategies are optimal. This is in line with the intuitive idea that when price impact is permanent, it should not matter how a trade is executed.  $\triangle$

**Example 2. (Exponential resilience)** Consider the continuous resilience function  $G(t) = e^{-\rho t}$  for  $\rho > 0$ . We will show that  $G(|t|) = e^{-\rho|t|}$  is the Fourier transform of the positive finite Borel measure

$$\mu(dy) = \frac{1}{\pi} \frac{\rho}{\rho^2 + y^2} dy.$$

Since  $G(|t|)$  and its transform are continuous and integrable, by the Fourier inversion theorem (cf. e.g. [31]),

$$G(|t|) = \int \frac{1}{\pi} \frac{\rho}{\rho^2 + y^2} e^{ity} dy \quad \text{if and only if} \quad \frac{1}{\pi} \frac{\rho}{\rho^2 + t^2} = \frac{1}{2\pi} \int G(|y|) e^{-ity} dy.$$

Using [33], we prove the equality to the right:

$$\begin{aligned}
\int G(|y|) e^{-ity} dy &= \int_{-\infty}^{\infty} e^{-\rho|y|} e^{-ity} dy \\
&= \int_{-\infty}^0 e^{\rho y} e^{-ity} dy + \int_0^{\infty} e^{-\rho y} e^{-ity} dy \\
&= \int_{-\infty}^0 e^{\rho y} (\cos(ty) - i \sin(ty)) dy + \int_0^{\infty} e^{-\rho y} (\cos(ty) - i \sin(ty)) dy \\
&= \int_0^{\infty} e^{-\rho u} (\cos(tu) + i \sin(tu)) du + \int_0^{\infty} e^{-\rho y} (\cos(ty) - i \sin(ty)) dy \\
&= 2 \int_0^{\infty} e^{-\rho y} \cos(ty) dy,
\end{aligned}$$

where the second last equality follows from the substitution  $u = -y$ . Using integration by parts, we derive

$$\begin{aligned}
\int_0^{\infty} e^{-\rho y} \cos(ty) dy &= \left[ \frac{1}{t} e^{-\rho y} \sin(ty) \right]_{y=0}^{\infty} + \int_0^{\infty} \frac{\rho}{t} e^{-\rho y} \sin(ty) dy \\
&= 0 + \frac{\rho}{t} \left( \left[ -\frac{1}{t} e^{-\rho y} \cos(ty) \right]_{y=0}^{\infty} - \int_0^{\infty} \frac{\rho}{t} e^{-\rho y} \cos(ty) dy \right) \\
&= \frac{\rho}{t} \left( 0 + \frac{1}{t} - \frac{\rho}{t} \int_0^{\infty} e^{-\rho y} \cos(ty) dy \right) \\
&= \frac{\rho}{t^2} - \frac{\rho^2}{t^2} \int_0^{\infty} e^{-\rho y} \cos(ty) dy.
\end{aligned}$$

Eliminating the integral expression, we find that

$$\left( 1 + \frac{\rho^2}{t^2} \right) \int_0^{\infty} e^{-\rho y} \cos(ty) dy = \frac{t^2 + \rho^2}{t^2} \int_0^{\infty} e^{-\rho y} \cos(ty) dy = \frac{\rho}{t^2},$$

and hence

$$\frac{1}{2\pi} \int G(|y|) e^{-ity} dy = \frac{2}{2\pi} \int_0^{\infty} e^{-\rho y} \cos(ty) dy = \frac{1}{\pi} \frac{t^2}{t^2 + \rho^2} \frac{\rho}{t^2} = \frac{1}{\pi} \frac{\rho}{t^2 + \rho^2}.$$

Now since the density function

$$\frac{1}{\pi} \frac{\rho}{t^2 + \rho^2} > 0 \text{ for all } t,$$

the support of  $\mu$  is given by  $\text{supp}(\mu) = \mathbb{R}$ , which clearly is not discrete. Therefore,  $G$  is positive definite and (2.6) gives the optimal order execution strategy, visualized in Figure 3.1 for  $\rho = 1$ . Here we used  $X_0 = 10$  and the equidistant time grid with starting point  $t_0 = 0$  and end point  $t_N = T = 10$ , both of which we will continue to use in all following examples in this chapter. We see that the optimal strategy exists of two identical, large trades at times 0 and  $T$ , and mutually identical, smaller trades in between those time points. The larger  $N$ , the smaller those trades become.  $\triangle$



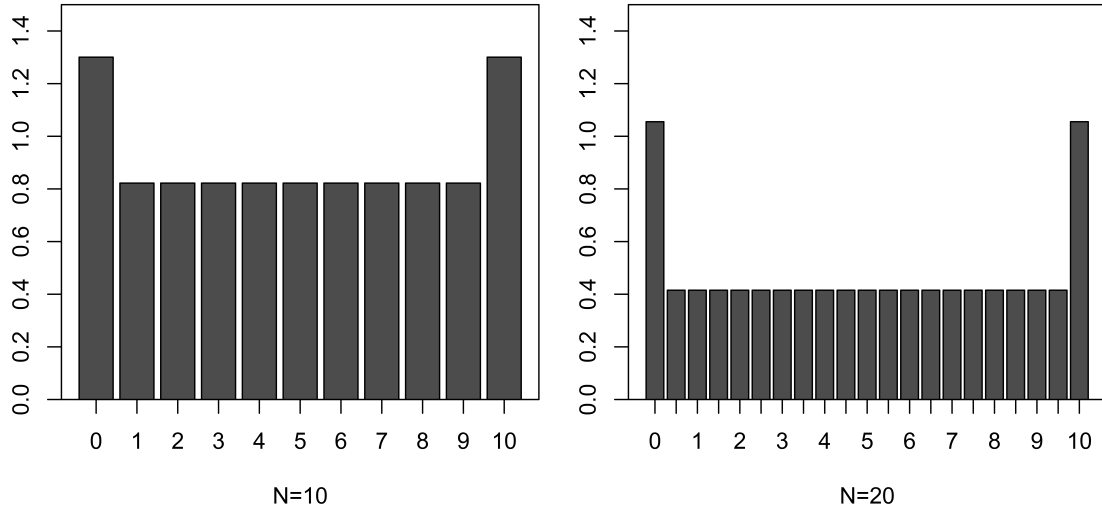


Figure 3.1: Optimal order execution strategies for the exponential resilience function  $G(t) = e^{-t}$ .

**Example 3. (Periodic resilience)** Consider the continuous resilience function

$$G(t) = \frac{e(e - \cos t)}{1 + e^2 - 2e \cos t},$$

shown in Figure 3.2 (note that  $G$  is not nonincreasing).

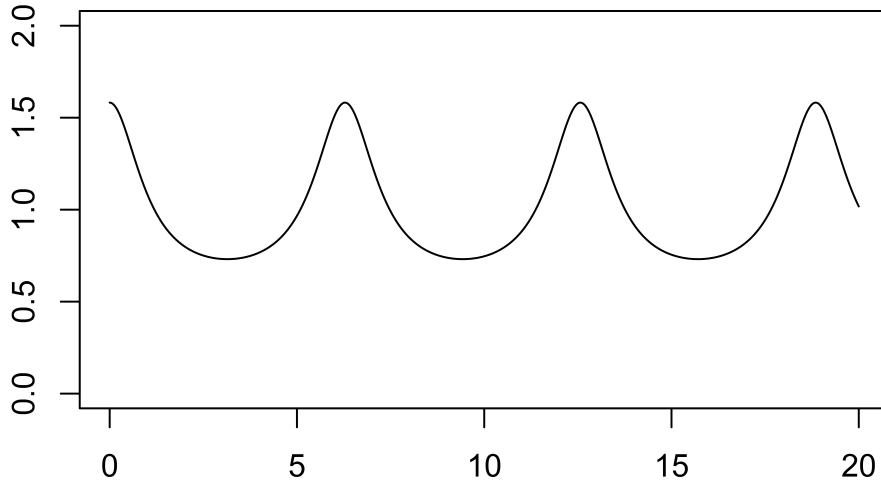


Figure 3.2: The periodic resilience function  $G(t) = \frac{e(e - \cos t)}{1 + e^2 - 2e \cos t}$ .

We will show that  $G(|t|) = G(t)$  is the Fourier transform of the purely discrete and finite measure

$$\mu = \frac{1}{2} \sum_{k=0}^{\infty} e^{-k} (\delta_k + \delta_{-k}),$$

where  $\delta_n$  is the Dirac measure

$$\delta_n(y) = \begin{cases} 1 & \text{if } y = n, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that since  $\mu$  is discrete,

$$\int e^{ity} \mu(dy) = \frac{1}{2} \sum_{k=0}^{\infty} e^{-k} (e^{itk} + e^{-itk}) = \frac{1}{2} \sum_{k=0}^{\infty} (e^{(it-1)k} + e^{(-it-1)k}).$$

From the well-known series

$$\sum_{k=0}^{N-1} r^k = \frac{1 - r^N}{1 - r},$$

we get that

$$\begin{aligned} \sum_{k=0}^{N-1} (e^{(it-1)k} + e^{(-it-1)k}) &= \sum_{k=0}^{N-1} ((e^{it-1})^k + (e^{-it-1})^k) \\ &= \frac{1 - e^{(it-1)N}}{1 - e^{it-1}} + \frac{1 - e^{(-it-1)N}}{1 - e^{-it-1}} \\ &= \frac{1 - e^{-N} (\cos(tN) + i \sin(tN))}{1 - e^{it-1}} \\ &\quad + \frac{1 - e^{-N} (\cos(tN) - i \sin(tN))}{1 - e^{-it-1}}. \end{aligned}$$

Taking the limit for  $N \rightarrow \infty$ , we end up with

$$\begin{aligned} \frac{1}{2} \sum_{k=0}^{\infty} (e^{(it-1)k} + e^{(-it-1)k}) &= \frac{1}{2} \left( \frac{1}{1 - e^{it-1}} + \frac{1}{1 - e^{-it-1}} \right) \\ &= \frac{2 - e^{it-1} - e^{-it-1}}{2(1 - e^{it-1})(1 - e^{-it-1})} \\ &= \frac{2 - e^{-1}(\cos t + i \sin t + \cos t - i \sin t)}{2(1 - e^{-1}(\cos t - i \sin t + \cos t + i \sin t) + e^{-2})} \\ &= \frac{1 - e^{-1} \cos t}{1 - 2e^{-1} \cos t + e^{-2}} \\ &= \frac{e(e - \cos t)}{1 + e^2 - 2e \cos t}. \end{aligned}$$

This proves that  $G$  is positive semidefinite. Since  $\text{supp}(\mu) = \mathbb{Z}$  is discrete, we cannot use Bochner's theorem and conclude that there exists a unique optimal strategy. However,  $M$  turns out to be invertible, and the strategy given by (2.6) is therefore optimal nevertheless (cf. Figure 3.3).  $\triangle$

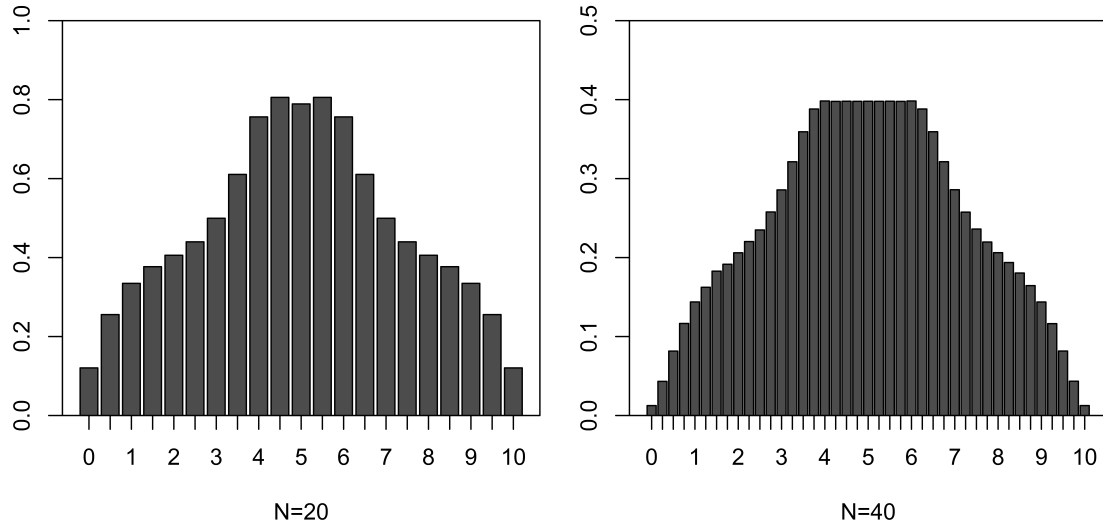


Figure 3.3: Optimal order execution strategies for the periodic resilience function  $G(t) = \frac{e(e - \cos t)}{1 + e^2 - 2e \cos t}$ .

### 3.2 Convex resilience functions

The examples above show that it can be quite a lot of work to verify whether a given resilience function is positive definite, using Bochner's theorem. The following proposition describes a large class of positive-definite functions by an easily verifiable characterization, thus simplifying matters.

**Proposition 3.** ([4]) *If the resilience function  $G$  is convex and nonconstant, then it is (strictly) positive definite.*

This means that in Example 2, we could have dropped the Fourier characterization, and simply used the fact that  $G(t) = e^{-\rho t}$  is convex and nonconstant.

*Proof.* In order to prove that a convex and nonconstant resilience function  $G$  is positive definite, we want to use Bochner's theorem. Therefore, we need to know whether  $G$  is continuous. As  $G(x)$  is convex, it is continuous except possibly in  $x = 0$  (cf. Theorem A on page 4 in [26]). Let  $\widetilde{M}$  denote the matrix corresponding to the continuous modification of  $G$ , i.e.

$$\widetilde{G}(x) = \begin{cases} G(x) & \text{for } x > 0, \\ \lim_{x \downarrow 0} G(x) & \text{for } x = 0. \end{cases}$$

On its diagonal,  $M$  has the value  $G(0)$ , while its continuous modification  $\widetilde{M}$  has the value  $G(0+)$  on its diagonal. We can thus write

$$M = (G(0) - G(0+))I_{N+1} + \widetilde{M}.$$

For discontinuous  $M$ , the term  $G(0) - G(0+)$  is positive, and can be explained by for instance execution costs. It follows that  $M$  is positive definite whenever  $\widetilde{M}$  is positive definite. This means that we can assume that  $G$  is continuous without losing generality, which enables us to use Bochner's theorem.

There is one thing we have to be careful about: the argument above only holds if the continuous modification of  $G$  is itself a convex and nonconstant function<sup>1</sup>. Whereas convexity is guaranteed for the continuous modification of any convex  $G$ , resilience functions of the form  $G(x) = a + b \cdot \mathbf{I}(x = 0)$  are convex and nonconstant, while their continuous modifications are constant functions. Here  $a \geq 0$  and  $b > 0$  are constants and  $\mathbf{I}(\cdot)$  denotes the indicator function. For resilience functions of this kind, the matrix  $M$  is given by

$$M = \begin{pmatrix} a+b & a & \cdots & a \\ a & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ a & \cdots & a & a+b \end{pmatrix} = b \cdot \mathbf{I}_{N+1} + \widetilde{M}, \quad (3.1)$$

where

$$\widetilde{M} = \begin{pmatrix} a & \cdots & a \\ \vdots & \ddots & \vdots \\ a & \cdots & a \end{pmatrix}$$

is the continuous modification of  $M$ . Let  $\mathbf{v} = (v_0, \dots, v_N)^T$  denote an eigenvector of  $\widetilde{M}$ . Then the equality  $\widetilde{M}\mathbf{v} = \lambda\mathbf{v}$  takes the form

$$\begin{pmatrix} a & \cdots & a \\ \vdots & \ddots & \vdots \\ a & \cdots & a \end{pmatrix} \begin{pmatrix} v_0 \\ \vdots \\ v_N \end{pmatrix} = a \left( \sum_{i=0}^N v_i \right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} v_0 \\ \vdots \\ v_N \end{pmatrix}.$$

It follows that either  $\lambda = 0$  or  $v_0 = \dots = v_N$ , i.e.  $\lambda = a(N+1) \geq 0$ . Similarly, let  $\mathbf{u} = (u_0, \dots, u_N)^T$  denote an eigenvector of  $M$ . Then by (3.1), the equality  $M\mathbf{u} = \nu\mathbf{u}$  takes the form

$$b \begin{pmatrix} u_0 \\ \vdots \\ u_N \end{pmatrix} + a \left( \sum_{i=0}^N u_i \right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \nu \begin{pmatrix} u_0 \\ \vdots \\ u_N \end{pmatrix},$$

or equivalently

$$a \left( \sum_{i=0}^N u_i \right) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = (\nu - b) \begin{pmatrix} u_0 \\ \vdots \\ u_N \end{pmatrix}.$$

It follows that either  $\nu = b > 0$  or  $u_0 = \dots = u_N$ , i.e.  $\nu = b + a(N+1) > 0$ . We find that all eigenvalues of  $M$  are positive, which implies that  $M$  is positive definite. This means that the assertion holds for resilience functions with constant continuous modifications as well.

We continue by showing that for  $G$  convex and nonconstant,  $G(|\cdot|)$  is the Fourier transform of a positive finite Borel measure. To ensure that the inverse transform exists, we introduce the function  $G_\varepsilon(x) := e^{-\varepsilon x} G(x)$  for  $\varepsilon > 0$ . Observe that like  $G$ ,  $G_\varepsilon$  is convex and decreasing. Since  $G$  is convex, it is absolutely continuous on  $(0, \infty)$  with existing right-hand derivative  $G'$  (cf. Theorem A and B on page 4-5 in [26]). The inverse Fourier

---

<sup>1</sup>The authors of [4] seem to jump over this.

transform of  $G_\varepsilon(|\cdot|)$  is given by

$$\begin{aligned} \frac{1}{2\pi} \int G_\varepsilon(|x|) e^{-ixz} \, dx &= \frac{1}{2\pi} \int_{-\infty}^0 G_\varepsilon(-x) e^{-ixz} \, dx + \frac{1}{2\pi} \int_0^\infty G_\varepsilon(x) e^{-ixz} \, dx \\ &= \frac{1}{2\pi} \int_0^\infty G_\varepsilon(x) (e^{ixz} + e^{-ixz}) \, dx \\ &= \frac{1}{\pi} \int_0^\infty G_\varepsilon(x) \cos(xz) \, dx. \end{aligned}$$

Now remember that  $G$  is nonincreasing and nonnegative, i.e.  $G(x) \in [0, G(0)]$  for all  $x \geq 0$ , and hence

$$G_\varepsilon(x) = e^{-\varepsilon x} G(x) \rightarrow 0 \text{ for } x \rightarrow \infty.$$

Integrating by parts, we find

$$\begin{aligned} \frac{1}{\pi} \int_0^\infty G_\varepsilon(x) \cos(xz) \, dx &= \frac{1}{\pi} \left[ G_\varepsilon(x) \frac{1}{z} \sin(xz) \right]_{x=0}^\infty - \frac{1}{\pi} \int_0^\infty G'_\varepsilon(x) \frac{1}{z} \sin(xz) \, dx \\ &= -\frac{1}{\pi} \int_{[0,\infty)} G'_\varepsilon(x) \, df_z(x), \end{aligned}$$

where

$$f_z(x) := -\frac{\cos(xz)}{z^2},$$

so that  $df_z(x) = f'_z(x) \, dx = \frac{1}{z} \sin(xz) \, dx$  is a signed Lebesgue-Stieltjes measure. The idea is to integrate by parts once more, so that we can use the nonnegativity of  $G''_\varepsilon$ , but we do not know whether  $G''_\varepsilon$  exists. However, since  $f_z$  is continuous and both  $f_z$  and  $G'_\varepsilon$  are of bounded variation, integration by parts for Lebesgue-Stieltjes integrals (cf. Theorem 3.36 and Exercise 3.34 in [15]) gives that<sup>2</sup>

$$\begin{aligned} \int_{[0,b]} G'_\varepsilon(x) \, df_z(x) &= G'_\varepsilon(b) f_z(b) - G'_\varepsilon(0) f_z(0) - \int_{[0,b]} f_z(x) \, dG'_\varepsilon(x) \\ &= G'_\varepsilon(b) f_z(b) + \frac{G'_\varepsilon(0)}{z^2} + \int_{[0,b]} \frac{\cos(xz)}{z^2} G''_\varepsilon(dx), \end{aligned}$$

where  $G''_\varepsilon(dx) = dG'_\varepsilon(x)$  is a Lebesgue-Stieltjes measure. Since

$$\int_{[0,\infty)} |G'_\varepsilon(x)| \, df_z(x) = - \int_{[0,\infty)} G'_\varepsilon(x) \, df_z(x) < \infty,$$

---

<sup>2</sup>From the notation used in [4], it is not clear whether the domain of integration includes 0. It is however important that it does, because when we for instance take  $G(x) = a + b \cdot \mathbf{1}(x=0)$  as in the start of this proof,  $G''(dx)$  will only be nonzero in 0. This we need in the last step of the proof.

it follows by dominated convergence (cf. Theorem 2.24 in [15]) that

$$\begin{aligned}
\int_{[0,\infty)} G'_\varepsilon(x) \, df_z(x) &= \int \lim_{b \rightarrow \infty} I_{[0,b]}(x) G'_\varepsilon(x) \, df_z(x) \\
&= \lim_{b \rightarrow \infty} \int I_{[0,b]}(x) G'_\varepsilon(x) \, df_z(x) \\
&= \lim_{b \rightarrow \infty} \int_{[0,b]} G'_\varepsilon(x) \, df_z(x) \\
&= \lim_{b \rightarrow \infty} \left( G'_\varepsilon(b) f_z(b) + \frac{G'_\varepsilon(0)}{z^2} + \int_{[0,b]} \frac{\cos(xz)}{z^2} G''_\varepsilon(dx) \right) \\
&= \frac{G'_\varepsilon(0)}{z^2} + \lim_{b \rightarrow \infty} \int I_{[0,b]}(x) \frac{\cos(xz)}{z^2} G''_\varepsilon(dx).
\end{aligned}$$

Now

$$\int_{[0,\infty)} \left| \frac{\cos(xz)}{z^2} \right| G''_\varepsilon(dx) \leq \int_{[0,\infty)} G''_\varepsilon(dx) = \lim_{x \rightarrow \infty} G'_\varepsilon(x) - G'_\varepsilon(0) = -G'_\varepsilon(0) < \infty, \quad (3.2)$$

so we can use the Lebesgue dominated convergence theorem once more and find that

$$\int_{[0,\infty)} G'_\varepsilon(x) \, df_z(x) = \frac{G'_\varepsilon(0)}{z^2} + \int_{[0,\infty)} \frac{\cos(xz)}{z^2} G''_\varepsilon(dx).$$

By (3.2),

$$-G'_\varepsilon(0) = \lim_{x \rightarrow \infty} G'_\varepsilon(x) - G'_\varepsilon(0) = \int_{[0,\infty)} G''_\varepsilon(dx),$$

which yields

$$\begin{aligned}
\frac{1}{2\pi} \int G_\varepsilon(|x|) e^{-ixz} \, dx &= \frac{1}{\pi} \int_0^\infty G_\varepsilon(x) \cos(xz) \, dx \\
&= -\frac{1}{\pi} \int_{[0,\infty)} G'_\varepsilon(x) \, df_z(x) \\
&= \frac{1}{\pi} \left( \int_{[0,\infty)} \frac{1}{z^2} G''_\varepsilon(dx) - \int_{[0,\infty)} \frac{\cos(xz)}{z^2} G''_\varepsilon(dx) \right) \\
&= \frac{1}{\pi} \int_{[0,\infty)} \frac{1 - \cos(xz)}{z^2} G''_\varepsilon(dx).
\end{aligned} \tag{3.3}$$

Since  $G$  is convex,  $G''_\varepsilon(dx)$  is a positive measure. Furthermore,

$$\frac{1 - \cos(xz)}{z^2} \geq 0,$$

with equality in only a countable number of points. Therefore, the last expression in (3.3), regarded as a function of  $z$ , is the density of a positive finite Borel measure  $\mu_\varepsilon$ , i.e.

$$\mu_\varepsilon([a,b]) := \int_a^b \frac{1}{2\pi} \int G_\varepsilon(|x|) e^{-ixz} \, dx \, dz = \int_a^b \frac{1}{\pi} \int_{[0,\infty)} \frac{1 - \cos(xz)}{z^2} G''_\varepsilon(dx) \, dz \geq 0$$

for  $0 < a < b$ . Equivalently,

$$G_\varepsilon(|z|) = \int e^{ixz} \mu_\varepsilon(dx),$$

and  $G_\varepsilon(|\cdot|)$  is the Fourier transform of a positive finite Borel measure  $\mu_\varepsilon$ . Hence, by Bochner's theorem,  $G_\varepsilon$  is positive semidefinite for all  $\varepsilon$ , i.e.

$$\sum_{i,j=0}^N x_i x_j e^{-\varepsilon|t_i-t_j|} G(|t_i-t_j|) \geq 0.$$

But this sum cannot become negative when taking the limit for  $\varepsilon$  decreasing to zero, and hence

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \sum_{i,j=0}^N x_i x_j e^{-\varepsilon|t_i-t_j|} G(|t_i-t_j|) &= \sum_{i,j=0}^N x_i x_j \left( \lim_{\varepsilon \downarrow 0} e^{-\varepsilon|t_i-t_j|} \right) G(|t_i-t_j|) \\ &= \sum_{i,j=0}^N x_i x_j G(|t_i-t_j|) \geq 0. \end{aligned}$$

This means that  $G$  is positive semidefinite. By Bochner's theorem, it remains to show that the support of  $\mu$  is not discrete, where  $\mu$  is the inverse Fourier transform of  $G$ . The continuity theorem (cf. Theorem 8.28 in [11]) says that for distribution functions  $F_n$  of random variables  $X_n$  with characteristic functions

$$\phi_n(z) = \mathbb{E}[e^{izX_n}] = \int e^{ixz} f_n(x) dx = \int e^{ixz} dF_n(x),$$

such that  $\lim_{n \rightarrow \infty} \phi_n(z) = h(z)$  exist for all  $z$  and  $h(z)$  is continuous in  $z = 0$ , there exists a distribution function  $F$  such that  $F_n \rightarrow F$  weakly. In addition,  $h$  is the characteristic function of  $F$ . In our case, since

$$G_\varepsilon(|z|) = \int e^{ixz} \mu_\varepsilon(dx),$$

and  $G_\varepsilon \rightarrow G$  pointwise for continuous  $G$ , the continuity theorem gives that  $\mu_\varepsilon$  converges weakly to  $\mu$  as a measure. By the portmanteau theorem (describing the equivalence of several definitions of weak convergence of measures, cf. Theorem 2.1 in [10]), this implies that

$$\mu([a, b]) \geq \limsup_{\varepsilon \downarrow 0} \mu_\varepsilon([a, b]) = \limsup_{\varepsilon \downarrow 0} \int_a^b \frac{1}{\pi} \int_{[0, \infty)} \frac{1 - \cos(xz)}{z^2} G'_\varepsilon(dx) dz.$$

Using that

$$g(x, z) := \frac{1 - \cos(xz)}{z^2}$$

is nonnegative and measurable as a continuous function, we can switch the integrals by the Fubini-Tonelli theorem (cf. Theorem 2.37 in [15]). Now since

$$G'_\varepsilon(x) = e^{-\varepsilon x} (G'(x) - \varepsilon G(x)) \rightarrow G'(x) \quad \text{for } \varepsilon \downarrow 0,$$

and

$$\int_a^b \frac{1 - \cos(xz)}{z^2} dz$$

is continuous and bounded as a function of  $x$ , we can use the portmanteau theorem once more, which yields

$$\frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz dG'_\varepsilon(x) \rightarrow \frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz dG'(x).$$

Equivalently,

$$\frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz G''_\varepsilon(dx) \rightarrow \frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz G''(dx),$$

which gives us that

$$\mu([a, b]) \geq \limsup_{\varepsilon \downarrow 0} \mu_\varepsilon([a, b]) \geq \frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz G''(dx).$$

But

$$\int_a^b \frac{1 - \cos(xz)}{z^2} dz > 0 \quad \text{for } 0 < a < b,$$

and by convexity of  $G$ , we have that  $G''_\varepsilon(dx) \geq 0$  with strict inequality in at least one point. Hence

$$\mu([a, b]) \geq \frac{1}{\pi} \int_{[0,\infty)} \int_a^b \frac{1 - \cos(xz)}{z^2} dz G''(dx) > 0$$

for all  $0 < a < b$ , and it follows that the support of  $\mu$  is not discrete. By Bochner's theorem, this means that  $G$  is strictly positive definite.  $\square$

Let us now take a look at some examples for which we can use Proposition 3.

**Example 4. (Power-law resilience)** According to [4] and [30], several empirical studies have found that market impact decays asymptotically with a power-law function of the form  $G(t) = \eta (1 + \lambda t)^{-\gamma}$  for some  $\eta, \gamma, \lambda > 0$  (cf. Figure 3.4).

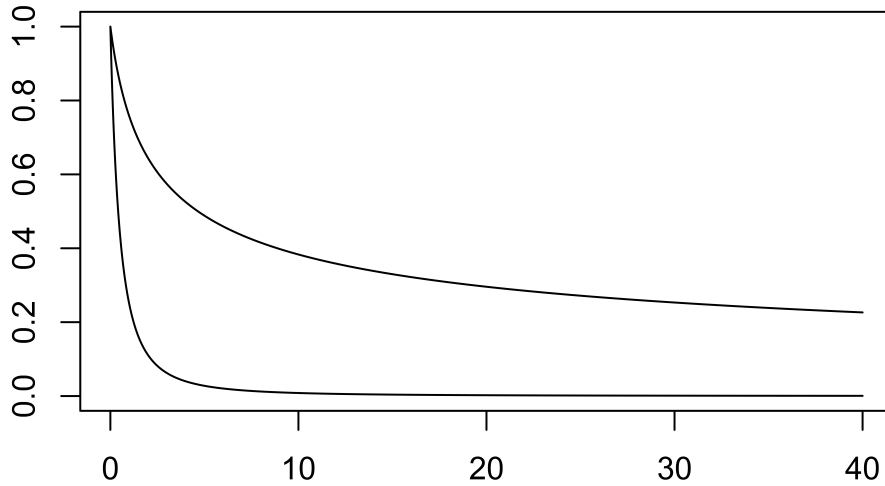


Figure 3.4: The power-law resilience function  $G(t) = (1 + t)^{-\gamma}$  for  $\gamma = 0.4$  (upper curve) and  $\gamma = 2$  (lower curve).

From Figure 3.4, clearly  $G$  is convex and nonconstant, and hence Proposition 3 applies. The unique optimal order execution strategy is shown in Figure 3.5 for different values of  $\gamma$ . We see that a higher value of  $\gamma$  forces the first and last trade to be smaller.  $\triangle$



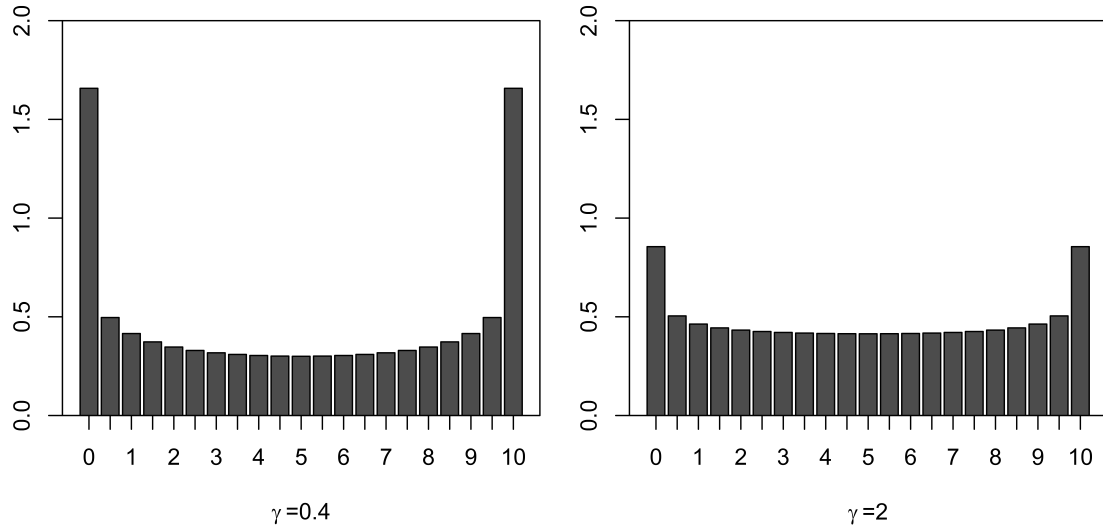


Figure 3.5: Optimal order execution strategies for the power-law resilience function  $G(t) = (1+t)^{-\gamma}$  with  $N = 20$ .

**Example 5. (Capped linear resilience)** In Figure 3.6, the resilience function  $G(t) = (1 - \rho t)^+$  with  $\rho > 0$  is shown.

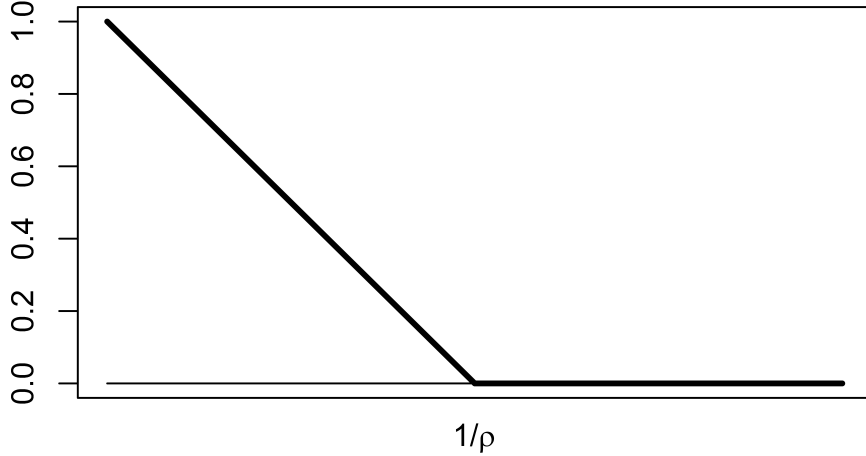


Figure 3.6: The capped linear resilience function  $G(t) = (1 - \rho t)^+$ .

Clearly,  $G$  is convex and nonconstant, and therefore positive definite. The optimal strategy depends on the size of  $\rho$ . If  $\rho \leq 1/T$ , then  $T \leq 1/\rho$ , and the effect of a trade at time 0 does not vanish before time  $T$ . If  $\rho > 1/T$ , then  $T > 1/\rho$ , and the effect of a trade at time  $t$  has vanished at time  $t + 1/\rho$ . Figure 3.7 shows the differences between the optimal strategies in the cases  $\rho \leq 1/T$  and  $\rho > 1/T$ . Note that there are a lot of trivial trades.

△

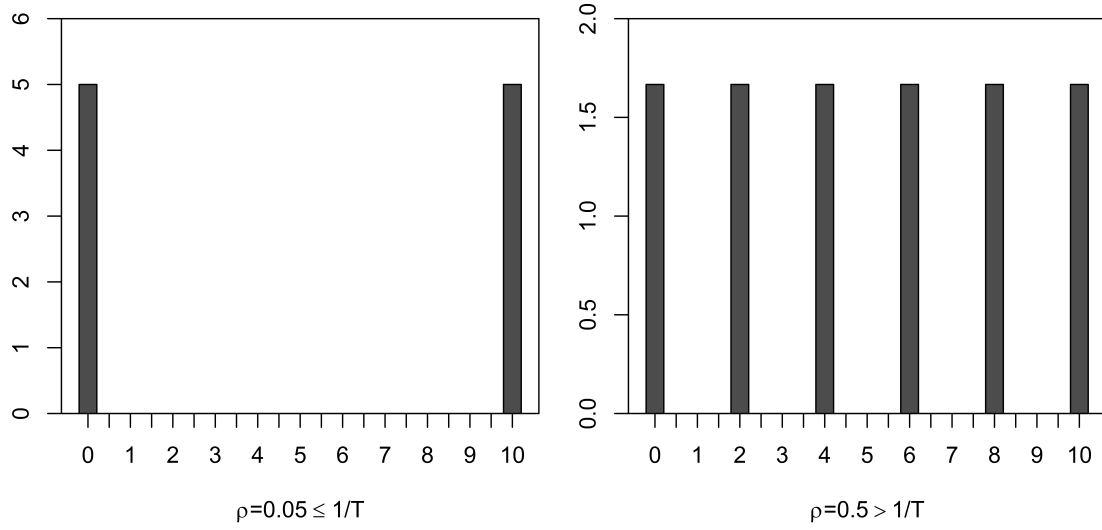


Figure 3.7: Optimal order execution strategies for the capped linear resilience function  $G(t) = (1 - \rho t)^+$  with  $N = 20$  and  $T = 10$ .

The following proposition formalizes what we see in Figure 3.7.

**Proposition 4.** ([4]) *Consider the case of capped linear resilience,  $G(t) = (1 - \rho t)^+$  for some  $\rho > 0$ .*

- (a) *When  $\rho \leq 1/T$ , the optimal strategy is independent of the underlying time grid. It consists of two symmetric trades of size  $X_0/2$  at  $t = 0$  and  $t = T$ , all other trades being zero.*
- (b) *Consider an equidistant grid of  $N + 1$  trading times, and suppose that  $\rho = k/T$ , where  $k \in \mathbb{N}$  divides  $N$ . Then the optimal strategy consists of  $k + 1$  equidistant trades of equal size.*

*Proof.* (a) Let  $\mathbf{x}^0 = (x_0^0, \dots, x_N^0)^T := (X_0/2, 0, 0, \dots, 0, X_0/2)^T$  denote the proposed strategy. Then

$$\begin{aligned}
 C(\mathbf{x}^0) &= \frac{1}{2} \sum_{i,j=0}^N x_i^0 x_j^0 G(|t_i - t_j|) \\
 &= \frac{1}{2} \left( \frac{X_0^2}{4} G(|t_0 - t_0|) + 2 \frac{X_0^2}{4} G(|t_N - t_0|) + \frac{X_0^2}{4} G(|t_N - t_N|) \right) \\
 &= \frac{X_0^2}{4} G(0) + \frac{X_0^2}{4} G(T) \\
 &= \frac{X_0^2}{4} (1 + (1 - \rho T)^+) = \frac{2 - \rho T}{4} X_0^2,
 \end{aligned}$$

since  $\rho \leq 1/T$ . Clearly,  $G$  is convex and nonconstant, which by Proposition 3 implies that  $G$  is strictly positive definite. We thus know from Theorem 1 that there exists a unique optimal strategy taking the form

$$\mathbf{x}^* = \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1}.$$

For the optimal strategy, the quadratic form  $C$  is thus explicitly given by (note that this relation holds in general)

$$\begin{aligned} C(\mathbf{x}^*) &= \frac{1}{2} (\mathbf{x}^*)^T M \mathbf{x}^* \\ &= \frac{1}{2} \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} \mathbf{1}^T M^{-1} M \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1} \\ &= \frac{1}{2} \frac{X_0^2}{\mathbf{1}^T M^{-1} \mathbf{1}}. \end{aligned}$$

We want to show that this equals

$$\frac{2 - \rho T}{4} X_0^2.$$

Let  $\mathbf{z} = (z_0, \dots, z_N) := M^{-1} \mathbf{1}$  and  $\Delta_i := t_i - t_{i-1}$ . Then the first and last component of the equation  $M\mathbf{z} = \mathbf{1}$  can be written as

$$\begin{aligned} \sum_{j=0}^N z_j G(|t_j - t_0|) &= \sum_{j=0}^N z_j \left( 1 - \rho \sum_{i=0}^j \Delta_i \right) = 1, \\ \sum_{j=0}^N z_j G(|t_j - t_N|) &= \sum_{j=0}^N z_j \left( 1 - \rho \sum_{i=j+1}^N \Delta_i \right) = 1. \end{aligned}$$

Adding these two equations together, we get

$$\sum_{j=0}^N z_j \left( 2 - \rho \sum_{i=0}^N \Delta_i \right) = \sum_{j=0}^N z_j (2 - \rho T) = 2,$$

and hence

$$\mathbf{1}^T M^{-1} \mathbf{1} = \mathbf{1}^T \mathbf{z} = \sum_{j=0}^N z_j = \frac{2}{2 - \rho T}.$$

Therefore,

$$C(\mathbf{x}^*) = \frac{1}{2} \frac{X_0^2}{\mathbf{1}^T M^{-1} \mathbf{1}} = \frac{2 - \rho T}{4} X_0^2.$$

By the uniqueness of  $\mathbf{x}^*$  (Theorem 1), it follows that  $\mathbf{x}^0$  is the optimal strategy.

(b) First, we adopt some notation. Let  $d = N/k$  be a positive integer, and define  $\Delta := T/N$ . We proceed as in (a). Let  $\mathbf{x}^0$  denote the proposed strategy, taking the value  $X_0/(k+1)$  at the trading times  $t_0, t_d, t_{2d}, \dots, t_N$ , and zero otherwise. Then

$$\begin{aligned} C(\mathbf{x}^0) &= \frac{1}{2} \sum_{i,j=0}^N x_i^0 x_j^0 G(|t_i - t_j|) \\ &= \frac{1}{2} \sum_{i,j=0}^k x_{di}^0 x_{dj}^0 G(|t_{di} - t_{dj}|) \\ &= \frac{1}{2} \sum_{i,j=0}^k \frac{X_0^2}{(k+1)^2} G(|t_{di} - t_{dj}|), \end{aligned}$$

since by the definition of  $\mathbf{x}^0$ ,  $x_{dm}^0 = X_0/(k+1)$  for  $m \in \{0, \dots, k\}$ , and  $x_l^0 = 0$  for  $l \notin \{0, d, 2d, \dots, N\}$ . We take a closer look at this expression:

$$\begin{aligned}
& \sum_{i,j=0}^k G(|t_{di} - t_{dj}|) \\
&= (1 - \rho \cdot 0)^+ + (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 2d\Delta)^+ + \dots + (1 - \rho \cdot kd\Delta)^+ & (i=0) \\
&+ (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 0)^+ + (1 - \rho \cdot d\Delta)^+ + \dots + (1 - \rho \cdot (k-1)d\Delta)^+ & (i=1) \\
&+ (1 - \rho \cdot 2d\Delta)^+ + (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 0)^+ + (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 2d\Delta)^+ \\
&\quad \quad \quad + \dots + (1 - \rho \cdot (k-2)d\Delta)^+ & (i=2) \\
&\quad \quad \quad \vdots & \vdots \\
&+ (1 - \rho \cdot (k-1)d\Delta)^+ + \dots + (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 0)^+ + (1 - \rho \cdot d\Delta)^+ & (i=k-1) \\
&+ (1 - \rho \cdot kd\Delta)^+ + \dots + (1 - \rho \cdot d\Delta)^+ + (1 - \rho \cdot 0)^+. & (i=k)
\end{aligned}$$

Now  $\rho d\Delta = k/T \cdot N/k \cdot T/N = 1$  and hence  $(1 - \rho \cdot md\Delta)^+ = 0$  for  $m \in \{1, \dots, k\}$ . This means that

$$\sum_{i,j=0}^k G(|t_{di} - t_{dj}|) = k+1,$$

from which we get that

$$C(\mathbf{x}^0) = \frac{X_0^2}{2(k+1)^2} \sum_{i,j=0}^k G(|t_{di} - t_{dj}|) = \frac{X_0^2}{2(k+1)}.$$

Now, as in (a), we want to show that for the optimal strategy  $\mathbf{x}^*$ ,

$$C(\mathbf{x}^*) = \frac{1}{2} \frac{X_0^2}{\mathbf{1}^T M^{-1} \mathbf{1}} = \frac{X_0^2}{2(k+1)}$$

holds. To this end, we again define  $\mathbf{z} = (z_0, \dots, z_N)^T := M^{-1} \mathbf{1}$ . This gives us the relation  $M\mathbf{z} = \mathbf{1}$ , i.e.  $\sum_{j=0}^N z_j G(|t_i - t_j|) = 1$  for all  $i \in \{0, 1, \dots, N\}$ . For  $i \in \{0, d, 2d, \dots, N\}$ ,

we get

$$\begin{aligned}
i = 0 : \quad 1 &= \sum_{j=0}^N z_j (1 - \rho \cdot j\Delta)^+ \\
&= z_0 + \sum_{j=1}^{d-1} z_j (1 - \rho \cdot j\Delta) \\
i = d : \quad 1 &= \sum_{j=0}^N z_j (1 - \rho \cdot |d - j|\Delta)^+ \\
&= \sum_{j=1}^{d-1} z_j (1 - \rho \cdot (d - j)\Delta) + z_d + \sum_{j=d+1}^{2d-1} z_j (1 - \rho \cdot (j - d)\Delta) \\
i = 2d : \quad 1 &= \sum_{j=0}^N z_j (1 - \rho \cdot |2d - j|\Delta)^+ \\
&= \sum_{j=d+1}^{2d-1} z_j (1 - \rho \cdot (2d - j)\Delta) + z_{2d} + \sum_{j=2d+1}^{3d-1} z_j (1 - \rho \cdot (j - 2d)\Delta) \\
&\quad \vdots \qquad \qquad \qquad \vdots \\
i = N - d : \quad 1 &= \sum_{j=0}^N z_j (1 - \rho \cdot |N - d - j|\Delta)^+ \\
&= \sum_{j=N-2d+1}^{N-d-1} z_j (1 - \rho \cdot (N - d - j)\Delta) + z_{N-d} \\
&\qquad \qquad \qquad + \sum_{j=N-d+1}^{N-1} z_j (1 - \rho \cdot (j - N + d)\Delta) \\
i = N : \quad 1 &= \sum_{j=0}^N z_j (1 - \rho \cdot |N - j|\Delta)^+ \\
&= \sum_{j=N-d+1}^{N-1} z_j (1 - \rho \cdot (N - d - j)\Delta) + z_N.
\end{aligned}$$

Now since for  $m \in \{0, \dots, k\}$ ,

$$\begin{aligned}
& \sum_{j=md+1}^{(m+1)d-1} z_j (1 - \rho \cdot (j - md)\Delta) + \sum_{j=md+1}^{(m+1)d-1} z_j (1 - \rho \cdot ((m+1)d - j)\Delta) \\
&= \sum_{j=md+1}^{(m+1)d-1} z_j (2 - \rho \cdot d\Delta) \\
&= \sum_{j=md+1}^{(m+1)d-1} z_j,
\end{aligned}$$

when taking the sum over all equations, we get

$$\begin{aligned}
k+1 &= z_0 + \sum_{j=1}^{d-1} z_j + z_d + \sum_{j=d+1}^{2d-1} z_j + z_{2d} + \dots + \sum_{j=N-d+1}^{N-1} z_j + z_N \\
&= \sum_{j=0}^N z_j = \mathbf{1}^T \mathbf{z} = \mathbf{1}^T M^{-1} \mathbf{1}.
\end{aligned}$$

Hence

$$C(\mathbf{x}^*) = \frac{1}{2} \frac{X_0^2}{\mathbf{1}^T M^{-1} \mathbf{1}} = \frac{X_0^2}{2(k+1)},$$

and  $\mathbf{x}^* = \mathbf{x}^0$  by uniqueness of the optimal strategy. □

### 3.3 Non-convex resilience functions

Recall the power-law resilience function

$$G(t) = \frac{1}{(1+t)^\gamma}.$$

In Example 4 we saw that the optimal strategy is well-behaved if we (for instance) take  $\gamma = 2$ . However, if we change the definition of power-law decay just a little, this may change drastically, as the following example shows.

**Example 6. (Alternative power-law resilience)** Consider the continuous resilience function

$$G(t) = \frac{1}{1+t^2},$$

which resembles the power-law resilience function

$$\bar{G}(t) = \frac{1}{(1+t)^2}.$$

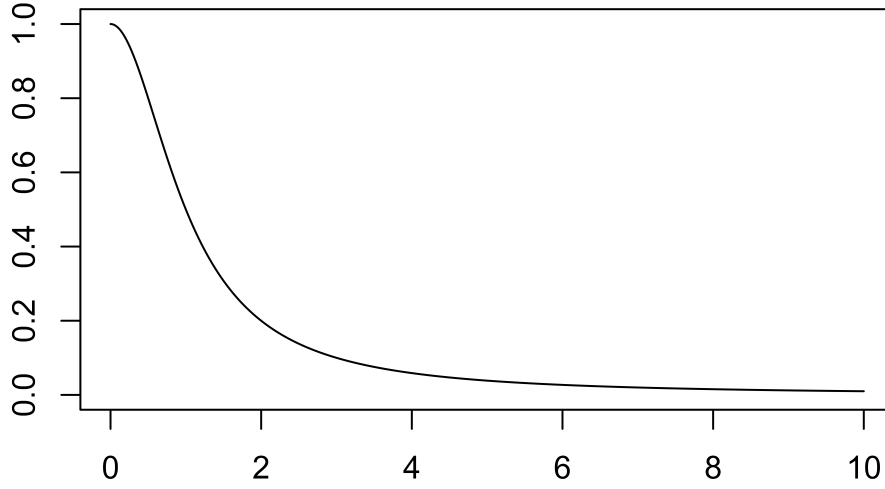


Figure 3.8: The alternative power-law resilience function  $G(t) = 1/(1+t^2)$ .

From Figure 3.8, we see that  $G$  is not convex. Therefore, we return to Bochner's theorem, and show that  $G(t) = G(|t|)$  is the Fourier transform of the measure

$$\mu(dy) = \frac{1}{2}e^{-|y|}dy.$$

We have that

$$\begin{aligned} \int e^{ity} \mu(dy) &= \int_{-\infty}^{\infty} \frac{1}{2} e^{ity} e^{-|y|} dy \\ &= \int_{-\infty}^0 \frac{1}{2} e^{ity} e^y dy + \int_0^{\infty} \frac{1}{2} e^{ity} e^{-y} dy \\ &= \int_{-\infty}^0 \frac{1}{2} e^y (\cos(ty) + i \sin(ty)) dy + \int_0^{\infty} \frac{1}{2} e^{-y} (\cos(ty) + i \sin(ty)) dy \\ &= \int_0^{\infty} \frac{1}{2} e^{-u} (\cos(tu) - i \sin(tu)) du + \int_0^{\infty} \frac{1}{2} e^{-y} (\cos(ty) + i \sin(ty)) dy \\ &= \int_0^{\infty} e^{-y} \cos(ty) dy. \end{aligned}$$

Now using Example 2 with  $\rho = 1$ , we find that

$$G(t) = \frac{1}{1+t^2}.$$

The support of  $\mu$ ,  $\text{supp}(\mu) = \mathbb{R}$ , is not discrete and hence  $G$  is positive definite by Bochner's theorem. The optimal order execution strategy is displayed in Figure 3.9.

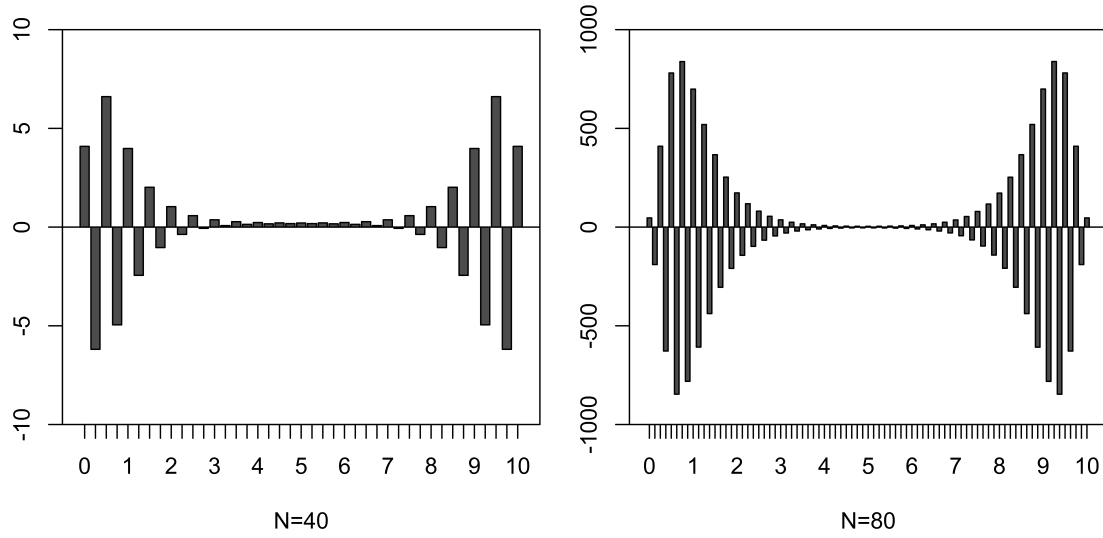


Figure 3.9: Optimal order execution strategies for the alternative power-law resilience function  $G(t) = 1/(1 + t^2)$ .

We see that even though  $X_0 = 10$  is strictly positive, negative trades occur. For large  $N$ , the optimal strategy oscillates strongly between selling and buying. In addition, the sizes of some of the child orders become larger and larger. It seems that we have found a new kind of price manipulation.  $\triangle$

The last example we consider, shows something similar.

**Example 7. (Gaussian resilience)** As we can see from Figure 3.10, the Gaussian resilience function  $G(t) = G(|t|) = \eta e^{-t^2}$  is not convex. Therefore, we compute its Fourier transform.

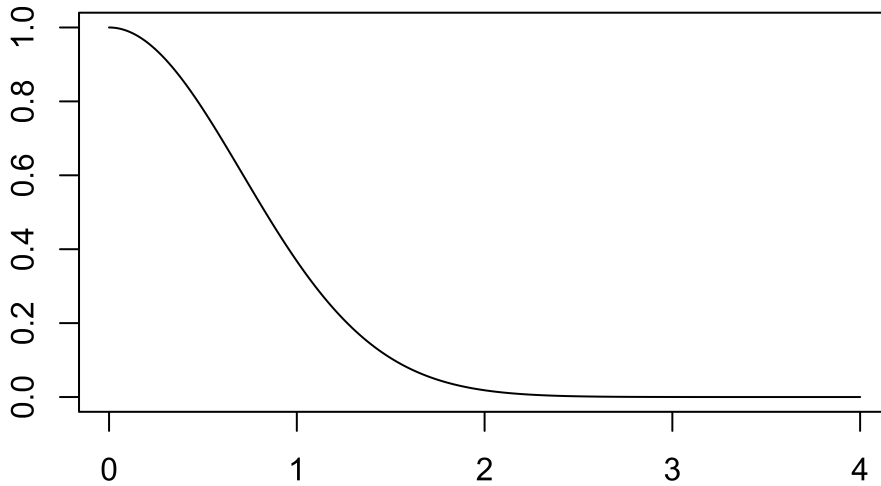


Figure 3.10: The Gaussian resilience function  $G(t) = e^{-t^2}$ .

First, note that

$$G'(t) = -2t\eta e^{-t^2} = -2t G(t). \quad (3.4)$$



Since  $\lim_{y \rightarrow \infty} G(y) = \lim_{y \rightarrow -\infty} G(y) = 0$ , the inverse Fourier transform of the left-hand side of this equation is given by

$$\begin{aligned} \mathcal{F}^{-1}(G'(t)) &= \frac{1}{2\pi} \int e^{-ity} G'(y) dy \\ &= \frac{1}{2\pi} \left[ e^{-ity} G(y) \right]_{y=-\infty}^{\infty} + \frac{1}{2\pi} \int it e^{-ity} G(y) dy \\ &= it \mathcal{F}^{-1}(G(t)), \end{aligned}$$

where we used Euler's formula. On the other hand, the inverse Fourier transform of the right-hand side of (3.4) is given by

$$\mathcal{F}^{-1}(-2t G(t)) = \frac{1}{2\pi} \int -2y e^{-ity} G(y) dy.$$

Using [32], we observe that (since  $e^{-ity} G(y) = \eta e^{-y(it+y)}$  and its derivative are continuous in both of their variables) by the Leibniz integral rule

$$\begin{aligned} \frac{d}{dt} \mathcal{F}^{-1}(G(t)) &= \frac{d}{dt} \left( \frac{1}{2\pi} \int e^{-ity} G(y) dy \right) \\ &= \frac{1}{2\pi} \int -iy e^{-ity} G(y) dy \\ &= \frac{i}{2} \mathcal{F}^{-1}(-2t G(t)), \end{aligned}$$

and hence taking the inverse Fourier transform on both sides of (3.4) yields

$$it \mathcal{F}^{-1}(G(t)) = \frac{2}{i} \left( \frac{d}{dt} \mathcal{F}^{-1}(G(t)) \right)$$

Writing  $\check{G}(t) := \mathcal{F}^{-1}(G(t))$ , we recognize the differential equation

$$\check{G}'(t) = -\frac{1}{2}t \check{G}(t),$$

which has the unique solution

$$\check{G}(t) = \check{G}(0) e^{-t^2/4},$$

with

$$\begin{aligned} \check{G}(0) &= \frac{1}{2\pi} \int e^{-iy \cdot 0} G(y) dy \\ &= \frac{1}{2\pi} \int \eta e^{-y^2} dy \\ &= \frac{\eta}{\sqrt{2\pi}} \int \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{2}y)^2/2} dy \\ &= \frac{\eta}{\sqrt{4\pi}} \int \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{\eta}{\sqrt{4\pi}}, \end{aligned}$$

where, in the last step, we used that the integral, as the distribution function of the standard normal distribution, integrates to one. We now have that

$$\check{G}(t) = \frac{1}{2\pi} \int e^{-ity} G(y) dy = \frac{\eta}{\sqrt{4\pi}} e^{-t^2/4}$$

and by the Fourier inversion theorem, this gives

$$\begin{aligned} G(t) &= \int e^{ity} \frac{\eta}{\sqrt{4\pi}} e^{-y^2/4} dy \\ &= \int e^{ity} \mu(dy), \end{aligned}$$

where

$$\mu(dy) = \frac{\eta}{\sqrt{4\pi}} e^{-y^2/4} dy$$

is a finite Borel measure with support  $\mathbb{R}$ , which is not discrete. Hence, by Bochner's theorem,  $G$  is positive definite, and the associated unique optimal trading strategy is displayed in Figure 3.11. We see that also in this case, the optimal strategy oscillates between buying and selling, and some of the child orders become disproportionately large. For  $N = 30$ , we see several buy and sell orders of more than 400 shares each, while the total order size is only  $X_0 = 10$ .  $\triangle$

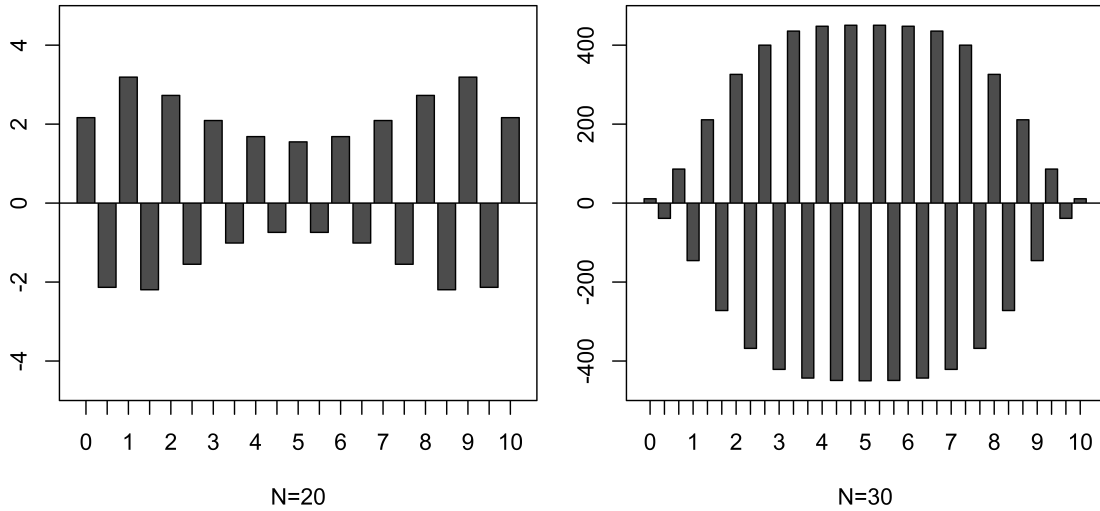


Figure 3.11: Optimal order execution strategies for the Gaussian resilience function  $G(t) = e^{-t^2}$ .

The last two examples show some behavior that reminds of price manipulation. Both of the corresponding resilience functions are not convex, and one could wonder whether this is a coincidence or not. In the next chapter, we will look deeper into this.

## Chapter 4

# Transaction-triggered price manipulation

In Chapter 2 we saw that requiring the resilience function to be positive definite, yielded the zero-trade strategy  $\xi = \mathbf{0}$  as the unique optimal strategy in the class of round trips, which excluded price manipulation. Nevertheless, we saw some examples in which the optimal strategy for trading a nontrivial amount of shares alternated between buying and selling. Clearly, models that admit this kind of behavior cannot be regarded viable and should be excluded. In fact, applying such alternating strategies may be illegal ([4]). Therefore, the following notion is introduced:

**Definition 1.** A market impact model admits *transaction-triggered price manipulation* when the expected execution costs of a sell (buy) program can be decreased by intermediate buy (sell) trades. More precisely, in our setting, the model admits transaction-triggered price manipulation if there exist  $X_0 \neq 0$ , a time grid  $0 \leq t_0 < \dots < t_N$  and  $\mathbf{y} \in \mathbb{R}^{N+1}$  such that  $\mathbf{y}^T \mathbf{1} = X_0$  and

$$C(\mathbf{y}) < \min \left\{ C(\mathbf{x}) \mid \mathbf{x}^T \mathbf{1} = X_0 \text{ and all components of } \mathbf{x} \in \mathbb{R}^{N+1} \text{ have the same sign} \right\}.$$

From this definition, we get that a model allows transaction-triggered price manipulation if there exists a buy-and-sell strategy that beats the optimal strategy for the case in which only buying or only selling is allowed. An immediate consequence of the definition, is given by the following corollary.

**Corollary 1.** ([4]) *The absence of transaction-triggered price manipulation implies the absence of standard price manipulation.*

*Proof.* We proceed by contraposition. Assume there exists a deterministic standard price manipulation strategy  $\xi$ , i.e.  $\xi^T \mathbf{1} = X_0 = 0$  and  $C(\xi) = \mathbb{E}[C(\xi)] = C(\xi) < 0$ . Then, if we use this strategy, but buy additional  $\varepsilon$  shares at time  $T$ , for  $\varepsilon > 0$  small enough,  $C(\xi_\varepsilon)$  will still be negative, by continuity of the quadratic form. Here  $\xi_\varepsilon$  denotes the new strategy  $\xi_\varepsilon = \xi + (0, \dots, 0, \varepsilon)^T$ , for which we have that  $\xi_\varepsilon^T \mathbf{1} = X_0 + \varepsilon = \varepsilon > 0$ . Since

$$\min \left\{ C(\mathbf{x}) \mid \mathbf{x}^T \mathbf{1} = \varepsilon \text{ and all components of } \mathbf{x} \in \mathbb{R}^{N+1} \text{ have the same sign} \right\} \geq 0,$$

we have found a transaction-triggered price manipulation strategy  $\xi_\varepsilon$ . The result follows by contraposition.  $\square$

## 4.1 Sufficiency of convexity

Our goal is now to formulate conditions which guarantee that all components of the optimal strategy  $\xi^*$  have the same sign. In Chapter 3 we saw that the resilience functions allowing transaction-triggered price manipulation were exactly those that were not convex. Indeed, we have the following theorem.

**Theorem 2.** ([4]) *For a convex and nonconstant resilience function  $G$  there are no transaction-triggered price manipulation strategies. If  $G$  is even strictly convex, then all trades in an optimal order execution strategy are strictly positive for a buy program and strictly negative for a sell program.*

In order to prove this theorem, we need the following lemma, in which all inequalities are to be understood componentwise.

**Lemma 1.** ([4]) *Let  $M$  be an invertible symmetric matrix.*

(a) *We have  $M^{-1}\mathbf{1} > \mathbf{0}$  or  $M^{-1}\mathbf{1} < \mathbf{0}$  if and only if there is no vector  $\mathbf{z}$  such that*

$$\mathbf{z}^T \mathbf{1} = 0, M\mathbf{z} \geq \mathbf{0}, M\mathbf{z} \neq \mathbf{0}. \quad (4.1)$$

(b) *We have  $M^{-1}\mathbf{1} \geq \mathbf{0}$  or  $M^{-1}\mathbf{1} \leq \mathbf{0}$  if and only if there is no vector  $\mathbf{z}$  such that*

$$\mathbf{z}^T \mathbf{1} = 0, M\mathbf{z} > \mathbf{0}.$$

*Proof.* (a) Suppose  $M^{-1}\mathbf{1} > \mathbf{0}$  or  $M^{-1}\mathbf{1} < \mathbf{0}$  and assume that there exists a vector  $\mathbf{z}$  such that  $\mathbf{z}^T \mathbf{1} = 0$ ,  $M\mathbf{z} \geq \mathbf{0}$ , and  $M\mathbf{z} \neq \mathbf{0}$ . Then since  $M^{-1}\mathbf{1} > \mathbf{0}$  or  $M^{-1}\mathbf{1} < \mathbf{0}$ , and  $M\mathbf{z} \geq \mathbf{0}$ ,  $M\mathbf{z} \neq \mathbf{0}$ , we get that  $(M\mathbf{z})^T M^{-1}\mathbf{1} > 0$  or  $(M\mathbf{z})^T M^{-1}\mathbf{1} < 0$ . On the other hand,

$$(M\mathbf{z})^T M^{-1}\mathbf{1} = \mathbf{z}^T M M^{-1}\mathbf{1} = \mathbf{z}^T \mathbf{1} = 0,$$

which is a contradiction.

Next, suppose that neither  $M^{-1}\mathbf{1} > \mathbf{0}$  nor  $M^{-1}\mathbf{1} < \mathbf{0}$ . Then  $\mathbf{x} := M^{-1}\mathbf{1}$  has at least one nonpositive and one nonnegative component. So  $\mathbf{x}$  has either a component  $x_l = 0$  or two components  $x_i < 0$  and  $x_j > 0$ . In the latter case, there exists a vector  $\mathbf{y}$  with  $y_i, y_j > 0$  and  $y_k = 0$  for all  $k \neq i, j$ , such that  $\mathbf{y}^T \mathbf{x} = 0$  (for instance by taking  $y_i = |x_j|$  and  $y_j = |x_i|$ ). For  $\mathbf{z} := M^{-1}\mathbf{y}$  we have that  $M\mathbf{z} = \mathbf{y} \geq \mathbf{0}$ ,  $M\mathbf{z} \neq \mathbf{0}$ , and

$$\mathbf{z}^T \mathbf{1} = \mathbf{y}^T M^{-1}\mathbf{1} = \mathbf{y}^T \mathbf{x} = 0.$$

This shows that in this case there exists a vector  $\mathbf{z}$  satisfying (4.1). In the other case, define  $\mathbf{y}$  by  $y_l = 1$  and  $y_k = 0$  for all  $k \neq l$ . Then by the same reasoning,  $\mathbf{z} := M^{-1}\mathbf{y}$  satisfies (4.1). This proves the assertion.

(b) We proceed as in part (a). Suppose that either  $M^{-1}\mathbf{1} \geq \mathbf{0}$  or  $M^{-1}\mathbf{1} \leq \mathbf{0}$ , and assume there exists a vector  $\mathbf{z}$  such that  $\mathbf{z}^T \mathbf{1} = 0$  and  $M\mathbf{z} > \mathbf{0}$ . Since  $M$  is invertible,  $M^{-1}\mathbf{1} \neq \mathbf{0}$ , and hence we get again that  $(M\mathbf{z})^T M^{-1}\mathbf{1} > 0$  or  $(M\mathbf{z})^T M^{-1}\mathbf{1} < 0$ , while on the other hand

$$(M\mathbf{z})^T M^{-1}\mathbf{1} = \mathbf{z}^T M M^{-1}\mathbf{1} = \mathbf{z}^T \mathbf{1} = 0.$$

This clearly is a contradiction.

Next, suppose that neither  $M^{-1}\mathbf{1} \geq \mathbf{0}$  nor  $M^{-1}\mathbf{1} \leq \mathbf{0}$ . Then  $\mathbf{x} := M^{-1}\mathbf{1}$  has at least one negative and one positive component, say  $x_i < 0$  and  $x_j > 0$ . Therefore, there exist  $\varepsilon > 0$

and a vector  $\mathbf{y}$  with  $y_i, y_j > 0$  and  $y_k = \varepsilon$  for all  $k \neq i, j$ , such that  $\mathbf{y}^T \mathbf{x} = 0$  (if the sum of all  $\varepsilon$ -terms is positive, this can be compensated by a high value for  $y_i$ ; if it is negative, it can be compensated by a high value for  $y_j$ ). But then, like in part (a),  $\mathbf{z} := M^{-1} \mathbf{y}$  satisfies  $M\mathbf{z} = \mathbf{y} > \mathbf{0}$  and

$$\mathbf{z}^T \mathbf{1} = \mathbf{y}^T M^{-1} \mathbf{1} = \mathbf{y}^T \mathbf{x} = 0,$$

which completes the proof.  $\square$

Using Lemma 1, we are now able to prove Theorem 2.

*Proof.* Assume  $G$  is convex and nonconstant, and recall that  $G$  is nonnegative and nonincreasing by definition. We want to show that  $\boldsymbol{\xi}^* \geq \mathbf{0}$  for a buy program and  $\boldsymbol{\xi}^* \leq \mathbf{0}$  for a sell program, with strict inequalities if  $G$  is strictly convex. As mentioned in the proof of Theorem 1,  $M$  is invertible. Clearly,  $M$  is also symmetric. If we can prove that  $M^{-1} \mathbf{1} \geq \mathbf{0}$  or  $M^{-1} \mathbf{1} \leq \mathbf{0}$ , we get (since  $M^{-1} \mathbf{1} \neq \mathbf{0}$ ) that  $\mathbf{1}^T M^{-1} \mathbf{1} > 0$  or  $\mathbf{1}^T M^{-1} \mathbf{1} < 0$  and hence, by (2.6), it follows that  $\boldsymbol{\xi}^* \geq \mathbf{0}$  if  $X_0 > 0$  and  $\boldsymbol{\xi}^* \leq \mathbf{0}$  if  $X_0 < 0$ , which is what we have to prove. If  $M^{-1} \mathbf{1} > \mathbf{0}$  or  $M^{-1} \mathbf{1} < \mathbf{0}$ , it follows in the same way that  $\boldsymbol{\xi}^* > \mathbf{0}$  if  $X_0 > 0$  and  $\boldsymbol{\xi}^* < \mathbf{0}$  if  $X_0 < 0$ .

By Lemma 1 (b), we want to show that there exists no vector  $\mathbf{z}$  such that  $\mathbf{z}^T \mathbf{1} = 0$  and  $M\mathbf{z} > \mathbf{0}$ . We proceed by induction on the dimension  $N$  of the time grid. Recall that  $\mathbf{z} = (z_0, \dots, z_N)^T \in \mathbb{R}^{N+1}$ . For  $N = 0$ ,  $\mathbf{z}^T \mathbf{1} = z_0$  and  $M\mathbf{z} = M_{00}z_0 = G(0)z_0$  (recall that  $M_{ij} = G(|t_i - t_j|)$ ). Clearly, there exists no  $\mathbf{z}$  satisfying  $\mathbf{z}^T \mathbf{1} = 0$  and  $M\mathbf{z} > \mathbf{0}$  in this case. Now suppose the same thing has already been proved for  $N = L - 1$ . We will show that if there exists a vector  $\mathbf{z}$  satisfying  $\mathbf{z}^T \mathbf{1} = 0$  and  $M\mathbf{z} > \mathbf{0}$  for  $N = L$ , then this must be true for the case  $N = L - 1$  as well. Since this violates the induction hypothesis, by contraposition, such a vector  $\mathbf{z}$  cannot exist for  $N = L$  either, which completes the proof by induction.

For the sake of clarity, we will in the sequel use the notation  $\mathbf{0}_N$  for the  $(N + 1)$ -dimensional zero vector, and  $\mathbf{1}_N$  for the  $(N + 1)$ -dimensional vector consisting of ones. Let us assume there exists a vector  $\mathbf{z}$  satisfying  $\mathbf{z}^T \mathbf{1}_L = 0$  and  $M\mathbf{z} > \mathbf{0}_L$ . Then, since  $G$  is nonnegative, there must be at least one  $k \in \{0, \dots, L\}$  such that  $z_k > 0$ . We consider three different cases:

If  $k = L$ , we have that

$$G(|t_L - t_m|) z_L \leq G(|t_{L-1} - t_m|) z_L \text{ for } m \in \{0, \dots, L-1\} \quad (4.2)$$

since  $G$  is nonincreasing. Let  $\tilde{M}$  be the  $(L, L)$ -dimensional matrix corresponding to the time grid  $\{t_0, \dots, t_{L-1}\}$ . Then, the  $L$ -dimensional vector  $\tilde{\mathbf{z}} := (z_0, \dots, z_{L-2}, z_{L-1} + z_L)^T$  satisfies both

$$\tilde{\mathbf{z}}^T \mathbf{1}_{L-1} = \mathbf{z}^T \mathbf{1}_L = 0$$

and

$$\begin{aligned}
\tilde{M}\tilde{\mathbf{z}} &= \begin{pmatrix} G(|t_0 - t_0|) & \cdots & G(|t_{L-1} - t_0|) \\ \vdots & \ddots & \vdots \\ G(|t_{L-1} - t_0|) & \cdots & G(|t_{L-1} - t_{L-1}|) \end{pmatrix} \begin{pmatrix} z_0 \\ \vdots \\ z_{L-2} \\ z_{L-1} + z_L \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=0}^{L-1} G(|t_j - t_0|) z_j + G(|t_{L-1} - t_0|) z_L \\ \vdots \\ \sum_{j=0}^{L-1} G(|t_j - t_{L-1}|) z_j + G(|t_{L-1} - t_{L-1}|) z_L \end{pmatrix} \\
&\geq \begin{pmatrix} \sum_{j=0}^L G(|t_j - t_0|) z_j \\ \vdots \\ \sum_{j=0}^L G(|t_j - t_{L-1}|) z_j \end{pmatrix},
\end{aligned} \tag{4.3}$$

where the inequality follows from (4.2). Now since all components of

$$M\mathbf{z} = \begin{pmatrix} \sum_{j=0}^L G(|t_j - t_0|) z_j \\ \vdots \\ \sum_{j=0}^L G(|t_j - t_L|) z_j \end{pmatrix}$$

are strictly positive by assumption, we get that  $\tilde{M}\tilde{\mathbf{z}} > \mathbf{0}_{L-1}$ , which is a violation of the induction hypothesis. This means that for  $k = L$ , we have proved that  $M^{-1}\mathbf{1} \geq 0$  or  $M^{-1}\mathbf{1} \leq 0$ .

We continue with the case  $k = 0$ . In line with the previous case, we have that

$$G(|t_m - t_0|) z_0 \leq G(|t_m - t_1|) z_0 \text{ for } m \in \{1, \dots, L\}. \tag{4.4}$$

Now let  $\hat{M}$  be the  $(L, L)$ -dimensional matrix corresponding to the time grid

$$\{t_1 - t_1, t_2 - t_1, \dots, t_L - t_1\}.$$

Then, the  $L$ -dimensional vector  $\hat{\mathbf{z}} := (z_0 + z_1, z_2, \dots, z_L)^T$  satisfies both

$$\hat{\mathbf{z}}^T \mathbf{1}_{L-1} = \mathbf{z}^T \mathbf{1}_L = 0$$

and

$$\begin{aligned}
\hat{M}\hat{\mathbf{z}} &= \begin{pmatrix} G(|t_1 - t_1|) & \cdots & G(|t_L - t_1|) \\ \vdots & \ddots & \vdots \\ G(|t_L - t_1|) & \cdots & G(|t_L - t_L|) \end{pmatrix} \begin{pmatrix} z_0 + z_1 \\ z_2 \\ \vdots \\ z_L \end{pmatrix} \\
&= \begin{pmatrix} G(|t_1 - t_1|) z_0 + \sum_{j=1}^L G(|t_j - t_1|) z_j \\ \vdots \\ G(|t_L - t_1|) z_0 + \sum_{j=1}^L G(|t_j - t_L|) z_j \end{pmatrix} \\
&\geq \begin{pmatrix} \sum_{j=0}^L G(|t_j - t_1|) z_j \\ \vdots \\ \sum_{j=0}^L G(|t_j - t_L|) z_j \end{pmatrix},
\end{aligned} \tag{4.5}$$

where the inequality follows from (4.4). Again, since all components of  $M\mathbf{z}$  are strictly positive by assumption, we get that  $\hat{M}\hat{\mathbf{z}} > \mathbf{0}_{L-1}$ , violating the induction hypothesis. This means that we have proved the assertion for  $k = 0$  as well.

Finally, suppose  $1 \leq k \leq L - 1$ . Let  $\alpha \in [0, 1]$  be such that  $t_k = \alpha t_{k-1} + (1 - \alpha)t_{k+1}$ , which is well-defined since  $k - 1 \geq 0$  and  $k + 1 \leq L$ . In this case, we have to use the fact that  $G$  is convex, meaning that for all  $s_1, s_2 \in [0, \infty)$  (also those  $s_1, s_2$  not being part of the time grid  $\{t_0, \dots, t_N\}$ ) and for all  $\alpha \in [0, 1]$ , we have that

$$G(\alpha s_1 + (1 - \alpha)s_2) \leq \alpha G(s_1) + (1 - \alpha)G(s_2).$$

First we observe that if  $t_k > t_l$  for  $0 \leq l \leq L$ , then  $t_{k-1} \geq t_l$ , and clearly  $t_{k+1} > t_l$ . If  $t_k < t_l$ , then  $t_{k+1} \leq t_l$ , and clearly  $t_{k-1} < t_l$ . Hence, for  $l \neq k$  (note that this implies that  $t_l \neq t_k$ ), we get that

$$\begin{aligned}
G(|t_k - t_l|) &= G(|\alpha t_{k-1} + (1 - \alpha)t_{k+1} - \alpha t_l - (1 - \alpha)t_l|) \\
&= G(|\alpha(t_{k-1} - t_l) + (1 - \alpha)(t_{k+1} - t_l)|) \\
&\leq \begin{cases} \alpha G(t_{k-1} - t_l) + (1 - \alpha)G(t_{k+1} - t_l) & \text{if } t_k > t_l \\ \alpha G(t_l - t_{k-1}) + (1 - \alpha)G(t_l - t_{k+1}) & \text{if } t_k < t_l \end{cases} \\
&= \alpha G(|t_{k-1} - t_l|) + (1 - \alpha)G(|t_{k+1} - t_l|).
\end{aligned}$$

But then clearly also

$$G(|t_k - t_l|) z_k \leq \alpha G(|t_{k-1} - t_l|) z_k + (1 - \alpha)G(|t_{k+1} - t_l|) z_k \text{ for } l \neq k \tag{4.6}$$

holds. Now the vector  $\bar{\mathbf{z}} := (z_0, \dots, z_{k-2}, z_{k-1} + \alpha z_k, z_{k+1} + (1 - \alpha)z_k, z_{k+2}, \dots, z_L)^T$  satisfies

$$\bar{\mathbf{z}}^T \mathbf{1}_{L-1} = z_0 + \dots + z_{k-1} + \alpha z_k + (1 - \alpha)z_k + z_{k+1} + \dots + z_L = \mathbf{z}^T \mathbf{1}_L = 0.$$

Letting  $\overline{M}$  the matrix corresponding to the time grid  $\{t_0, \dots, t_{k-1}, t_{k+1}, \dots, t_L\}$ , i.e.

$$\overline{M} = \begin{pmatrix} G(|t_0 - t_0|) & \cdots & G(|t_0 - t_{k-1}|) & G(|t_0 - t_{k+1}|) & \cdots & G(|t_0 - t_L|) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ G(|t_{k-1} - t_0|) & \cdots & G(|t_{k-1} - t_{k-1}|) & G(|t_{k-1} - t_{k+1}|) & \cdots & G(|t_{k-1} - t_L|) \\ G(|t_{k+1} - t_0|) & \cdots & G(|t_{k+1} - t_{k-1}|) & G(|t_{k+1} - t_{k+1}|) & \cdots & G(|t_{k+1} - t_L|) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ G(|t_L - t_0|) & \cdots & G(|t_L - t_{k-1}|) & G(|t_L - t_{k+1}|) & \cdots & G(|t_L - t_L|) \end{pmatrix},$$

we also have that  $\overline{M}\mathbf{z} =$

$$\begin{pmatrix} \sum_{j=0, j \neq k}^L G(|t_0 - t_j|) z_j + \alpha G(|t_0 - t_{k-1}|) z_k + (1 - \alpha) G(|t_0 - t_{k+1}|) z_k \\ \vdots \\ \sum_{j=0, j \neq k}^L G(|t_{k-1} - t_j|) z_j + \alpha G(|t_{k-1} - t_{k-1}|) z_k + (1 - \alpha) G(|t_{k-1} - t_{k+1}|) z_k \\ \sum_{j=0, j \neq k}^L G(|t_{k+1} - t_j|) z_j + \alpha G(|t_{k+1} - t_{k-1}|) z_k + (1 - \alpha) G(|t_{k+1} - t_{k+1}|) z_k \\ \vdots \\ \sum_{j=0, j \neq k}^L G(|t_L - t_j|) z_j + \alpha G(|t_L - t_{k-1}|) z_k + (1 - \alpha) G(|t_L - t_{k+1}|) z_k \end{pmatrix} \geq \begin{pmatrix} \sum_{j=0}^L G(|t_0 - t_j|) z_j \\ \vdots \\ \sum_{j=0}^L G(|t_{k-1} - t_j|) z_j \\ \sum_{j=0}^L G(|t_{k+1} - t_j|) z_j \\ \vdots \\ \sum_{j=0}^L G(|t_L - t_j|) z_j \end{pmatrix}, \quad (4.7)$$

where the inequality follows from (4.6). Again using the fact that all components of  $M\mathbf{z}$  are strictly positive by assumption, we find also in this case that  $\overline{M}\mathbf{z} > \mathbf{0}_{L-1}$ , which is impossible by the induction hypothesis. This finishes the first part of the proof.

The second part of the proof is parallel to the first part. By Lemma 1 (a), we want to show that there exist no vector  $\mathbf{z}$  such that

$$\mathbf{z}^T \mathbf{1} = 0, \quad M\mathbf{z} \geq \mathbf{0}, \quad M\mathbf{z} \neq \mathbf{0}.$$

For  $N = 0$ , this is trivial. Assume inductively that for the case  $N = L$  there exists a vector  $\mathbf{z}$  satisfying (4.1). Now remark that if  $G$  is strictly convex, it has to be strictly



decreasing as well, and the inequalities in (4.2), (4.4) and (4.6) become strict. This gives strict inequalities in (4.3), (4.5) and (4.7) as well. Now since  $M\mathbf{z} \geq \mathbf{0}$  by assumption, in all three cases we find a vector  $\mathbf{z}$  satisfying (4.1), thus violating the induction hypothesis. This finishes the proof.

As a last remark, it might be worthwhile noting that without strict convexity, we would only have been able to show that there exists a vector  $\bar{\mathbf{z}} \in \mathbb{R}^L$  satisfying  $\bar{\mathbf{z}}^T \mathbf{1}_{L-1} = 0$  and  $\bar{M}\bar{\mathbf{z}} \geq \mathbf{0}_{L-1}$ , but not that  $\bar{M}\bar{\mathbf{z}} \neq \mathbf{0}_{L-1}$ . Here  $\bar{M}$  denotes the concerning  $L \times L$  matrix. Indeed, if for  $\mathbf{z} \in \mathbb{R}^{L+1}$ ,  $M\mathbf{z} = (0, \dots, 0, a_k, 0, \dots, 0)^T$  with  $a_k > 0$  on place  $k$ , then  $M\mathbf{z} \geq \mathbf{0}$ ,  $M\mathbf{z} \neq \mathbf{0}$ , but still  $\bar{M}\bar{\mathbf{z}} = \mathbf{0}$ .  $\square$

## 4.2 Necessity of convexity

By Theorem 2, convexity of  $G$  is sufficient to exclude transaction-triggered price manipulation. The following proposition is a partial converse to Theorem 2. It says that for all time grids, the strongest decrease of  $G$  has to take place at the beginning (cf. Figure 4.1).

**Proposition 5.** ([4]) *Suppose that  $G$  is positive definite and that there exist  $s, t > 0$ ,  $s \neq t$  such that*

$$G(0) - G(s) < G(t) - G(t + s). \quad (4.8)$$

*Then the model admits transaction-triggered price manipulation strategies.*

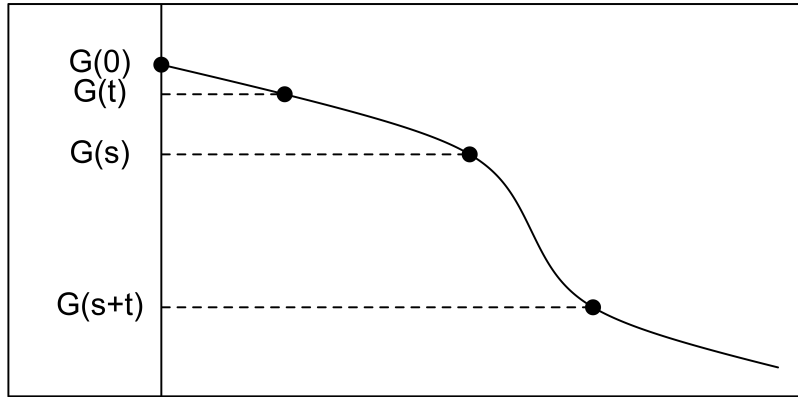


Figure 4.1: Visualization of Proposition 5.

*Proof.* We consider the case  $N = 2$  with the time grid  $t_0 = 0$ ,  $t_1 = s$  and  $t_2 = t + s$ . In this case,  $M$  is given by

$$M = \begin{pmatrix} G(0) & G(s) & G(t+s) \\ G(s) & G(0) & G(t) \\ G(t+s) & G(t) & G(0) \end{pmatrix}.$$

We will try to construct a transaction-triggered price manipulation strategy. Since  $G$  is (strictly) positive definite, we get from Theorem 1 that the optimal strategy is a multiple

of  $M^{-1}\mathbf{1}$ . In our case,  $M^{-1}$  follows from the equation  $(\det M) M^{-1} =$

$$\begin{pmatrix} G(0)^2 - G(t)^2 & -G(0)G(s) + G(t)G(t+s) & G(s)G(t) - G(0)G(t+s) \\ -G(0)G(s) + G(t)G(t+s) & G(0)^2 - G(t+s)^2 & -G(0)G(t) + G(s)G(t+s) \\ G(s)G(t) - G(0)G(t+s) & -G(0)G(t) + G(s)G(t+s) & G(0)^2 - G(s)^2 \end{pmatrix}$$

Now if the optimal strategy is a multiple of  $M^{-1}\mathbf{1}$ , then it is also a multiple of

$$\mathbf{z} = (z_0, z_1, z_2) := (\det M)M^{-1}\mathbf{1}$$

which we can compute directly. Since  $G$  is nonincreasing, we have that

$$\begin{aligned} z_0 &= G(0)^2 - G(t)^2 - G(0)G(s) + G(t)G(t+s) + G(s)G(t) - G(0)G(t+s) \\ &= (G(0) - G(t))(G(0) + G(t) - G(s) - G(t+s)) \\ &= (G(0) - G(t))((G(0) - G(s)) + (G(t) - G(t+s))) \geq 0, \end{aligned}$$

as all terms become nonnegative. Also

$$\begin{aligned} z_2 &= G(s)G(t) - G(0)G(t+s) - G(0)G(t) + G(s)G(t+s) + G(0)^2 - G(s)^2 \\ &= (G(0) - G(s))(G(0) + G(s) - G(t) - G(t+s)) \\ &= (G(0) - G(s))((G(0) - G(t)) + (G(s) - G(t+s))) \geq 0, \end{aligned}$$

and hence in an optimal strategy, nothing will be sold at the beginning or end of the trading period. Assets may be bought, or the trader stays inactive. From this, it follows that  $z_1 < 0$  when dealing with a sell program. When dealing with a buy program, we should have  $z_1 \geq 0$ , for otherwise the optimal strategy is a transaction-triggered price manipulation strategy. However, using (4.8),

$$\begin{aligned} z_1 &= -G(0)G(s) + G(t)G(t+s) + G(0)^2 - G(t+s)^2 - G(0)G(t) + G(s)G(t+s) \\ &= (G(0) - G(t+s))(G(0) + G(t+s) - G(s) - G(t)) \\ &= (G(0) - G(t+s))((G(0) - G(s)) - (G(t) - G(t+s))) < 0, \end{aligned}$$

independent of  $X_0$ , i.e. also when dealing with a buy program. Hence there exist  $X_0 \neq 0$  (namely all  $X_0 > 0$ ), a time grid  $0 = t_0 < t_1 < t_2$  and  $\mathbf{z} \in \mathbb{R}^3$  such that  $\mathbf{z}^T \mathbf{1} = X_0$  and

$$C(\mathbf{z}) < \min \left\{ C(\mathbf{x}) \mid \mathbf{x}^T \mathbf{1} = X_0 \text{ and all components of } \mathbf{x} \in \mathbb{R}^{N+1} \text{ have the same sign} \right\}.$$

The model thus admits transaction-triggered price manipulation strategies.  $\square$

From Figure 4.1, it is clear that condition (4.8) corresponds to a violation of convexity in a neighborhood of zero. This means that Theorem 2 and Proposition 5 together designate the nonnegative, nonincreasing, nonconstant, convex functions as the only resilience functions for which the corresponding market impact model is well-behaved. This can be

regarded as the main result concerning second-generation linear market impact models in discrete time.

The concavity around  $t = 0$  for resilience functions as the one displayed in Figure 4.1, can economically be interpreted as a delayed reaction of the market in response to a price shock ([4]). Markets described by such resilience functions are less efficient than markets described by convex resilience functions, and transaction-triggered price manipulation strategies exploit this market inefficiency.

One last remark on transaction-triggered price manipulation: Oscillatory effects as in Example 6 and 7 might seem mathematical irregularities that do not occur in reality. However, Alfonsi et al. (2012, [4]) point out that there are indications that through interaction of the trading algorithms of high-frequency traders (HFTs), such effects in fact can appear in reality. We quote from [14]: “Still lacking sufficient demand from fundamental buyers or cross-market arbitrageurs, HFTs began to quickly buy and then resell contracts to each other generating a “hot-potato” volume effect as the same positions were rapidly passed back and forth. Between 2:45:13 and 2:45:27, HFTs traded over 27,000 contracts, which accounted for about 49 percent of the total trading volume, while buying only about 200 additional contracts net.”



## Chapter 5

# Some properties of optimal strategies

In this chapter, we take a closer look at some qualitative properties of the optimal strategies given by Theorem 1. Looking back at the examples in Chapter 3, we see that the optimal strategies are symmetric, that is, time reversible. We have the following proposition formalizing this.

**Proposition 6.** ([4]) *Suppose that  $G$  is positive definite and that the time grid is reversible, i.e.  $t_i = t_N - t_{N-i}$  for  $i \in \{0, \dots, N\}$ . Then the optimal strategy  $\boldsymbol{\xi}^* = (\xi_{t_0}^*, \dots, \xi_{t_N}^*)$  is time reversible in the sense that  $\xi_{t_i}^* = \xi_{t_{N-i}}^*$  for  $i \in \{0, \dots, N\}$ .*

*Proof.* Let  $\hat{\mathbf{x}}$  denote the time reversal of the optimal strategy  $\mathbf{x}^*$ , i.e.  $\hat{x}_i = x_{N-i}^*$  for  $i \in \{0, \dots, N\}$ . Then

$$\begin{aligned}
 C(\hat{\mathbf{x}}) &= \frac{1}{2} \sum_{i,j=0}^N \hat{x}_i \hat{x}_j G(|t_i - t_j|) \\
 &= \frac{1}{2} \sum_{i,j=0}^N x_{N-i}^* x_{N-j}^* G(|t_N - t_{N-i} - (t_N - t_{N-j})|) \\
 &= \frac{1}{2} \sum_{i,j=0}^N x_{N-i}^* x_{N-j}^* G(|t_{N-j} - t_{N-i}|) \\
 &= \frac{1}{2} \sum_{i,j=0}^N x_i^* x_j^* G(|t_j - t_i|) = C(\mathbf{x}^*).
 \end{aligned} \tag{5.1}$$

Since, by assumption,  $G$  is positive definite, we get from Theorem 1 that the optimal strategy is unique<sup>1</sup>. It thus follows that  $\hat{\mathbf{x}} = \mathbf{x}^*$ , i.e.  $x_i^* = x_{N-i}^*$  for all  $i \in \{0, \dots, N\}$ .  $\square$

Next, we take a look at the signs of the components of  $\boldsymbol{\xi}^*$ . Looking back at Example 6 and 7, we see that in both cases, the first and the last trade have the same sign as  $X_0$ , even though many of the child orders have opposite sign. This is more generally true:

---

<sup>1</sup>The authors of [4] seem to forget that this can be used.

**Proposition 7.** ([4]) *Suppose that  $G$  is (strictly) positive definite and that the time grid is equidistant, i.e.  $t_n = nT/N$ . Then for  $X_0 \neq 0$ , the first and last trades of the optimal strategy are nonzero and have the same sign as  $X_0$ .*

*Proof.* Equation (2.6) suggests to look at  $M^{-1}\mathbf{1}$ . To be able to say something about the inverse of  $M$ , we will use that  $M$  is a Toeplitz matrix, i.e. a matrix of the form

$$\begin{pmatrix} T_0 & T_1 & \cdots & T_{D-1} & T_D \\ T_{-1} & T_0 & \ddots & & T_{D-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ T_{1-D} & & \ddots & T_0 & T_1 \\ T_{-D} & T_{1-D} & \cdots & T_{-1} & T_0 \end{pmatrix}.$$

This follows immediately from the fact that the time grid is equidistant, and it enables us to use an algorithm by S. Barnett (1990) for the computation of the inverse of a Toeplitz matrix, described in [8]. For this purpose, we need some notation: we write the  $N \times N$  matrix  $M = M_N$  as

$$M_N = \begin{pmatrix} m_0 & \mathbf{r}_{N-1}^T \\ \mathbf{r}_{N-1} & M_{N-1} \end{pmatrix}, \quad (5.2)$$

where  $m_0 = G(0) > 0$  (because  $G$  is strictly positive definite), and  $\mathbf{r}_{N-1}$  is the  $(N-1)$ -dimensional column vector consisting of the corresponding elements of  $M_N$ . The matrix  $M_{N-1}$  is again a (symmetric) Toeplitz matrix, consisting of all elements of  $M_N$  except the first row and column. Using the same notation, the matrix  $Y_N := M_N^{-1}$  will be written as

$$Y_N = \begin{pmatrix} y_{N-1} & \mathbf{f}_{N-1}^T \\ \mathbf{f}_{N-1} & Y_{N-1} \end{pmatrix}.$$

Here  $y_{N-1} \in \mathbb{R}$ ,  $\mathbf{f}_{N-1}$  is an  $(N-1)$ -dimensional column vector, and  $Y_{N-1}$  is an  $(N-1) \times (N-1)$  matrix. We need some more notation:  $K$  will denote a reverse unit matrix of appropriate dimension, i.e.

$$K = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Furthermore,  $r'_{n+1}$  denotes the last component of the vector  $\mathbf{r}_{n+1}$ , i.e.  $\mathbf{r}_{n+1} = (\mathbf{r}_n, r'_{n+1})$ . Finally,  $\delta_n := y_n r'_{n+1} + \mathbf{f}_n^T K \mathbf{f}_n$ . Using the starting points  $y_0 = 1/m_0$  and  $\mathbf{f}_0 = 0$ , we have the following recursion from [8]:

$$y_{n+1} = \frac{y_n}{1 - \delta_n^2}, \quad \mathbf{f}_{n+1} = \frac{1}{1 - \delta_n^2} \begin{pmatrix} \mathbf{f}_n - \delta_n K \mathbf{f}_n \\ -\delta_n y_n \end{pmatrix}. \quad (5.3)$$

In addition, [8] gives us that

$$\det M_n = \frac{1}{y_0 y_1 \cdots y_{n-1}}. \quad (5.4)$$

Sylvester's criterion states that a matrix is positive definite if and only if all its leading principle minors are positive, i.e. if and only if all upper left  $n \times n$  corners of  $M$  have

positive determinant for  $n \in \{1, \dots, N\}$  (cf. [23]). Since  $M = M_N$  is a Toeplitz matrix, the upper left  $n \times n$  corner of  $M$  is identical to the lower right  $n \times n$  corner for all  $n \in \{1, \dots, N\}$ . This means that by (5.2), the matrix  $M_n$  has a positive determinant for every  $n \in \{1, \dots, N\}$ . It follows from (5.4) that  $y_n > 0$  for all  $n$ , and the recursion (5.3) yields

$$\delta_n^2 < 1. \quad (5.5)$$

By Theorem 1 and Proposition 6, the first and last trades are both equal to the first component of

$$\frac{X_0}{\mathbf{1}_N^T Y_N \mathbf{1}_N} Y_N \mathbf{1}_N.$$

In order for the first and last trades to be nonzero and have the same sign as  $X_0$ , we should have that

$$\frac{(Y_N \mathbf{1}_N)_1}{\mathbf{1}_N^T Y_N \mathbf{1}_N} > 0,$$

where

$$(Y_N \mathbf{1}_N)_1 = y_{N-1} + \mathbf{f}_{N-1}^T \mathbf{1}_{N-1}$$

denotes the first component of  $Y_N \mathbf{1}_N$ . As  $\mathbf{1}^T M^{-1} \mathbf{1} > 0$  (cf. proof of Theorem 1), we show by induction on  $n$  that  $y_n + \mathbf{f}_n^T \mathbf{1}_n$  is strictly positive. For  $n = 0$ , the assertion clearly holds:

$$y_0 + \mathbf{f}_0^T \mathbf{1}_0 = \frac{1}{m_0} > 0.$$

Now by the recursion (5.3),

$$\begin{aligned} y_{n+1} + \mathbf{f}_{n+1}^T \mathbf{1}_{n+1} &= \frac{y_n}{1 - \delta_n^2} + \frac{1}{1 - \delta_n^2} \begin{pmatrix} \mathbf{f}_n - \delta_n K \mathbf{f}_n \\ -\delta_n y_n \end{pmatrix}^T \mathbf{1}_{n+1} \\ &= \frac{1}{1 - \delta_n^2} (y_n + (\mathbf{f}_n^T - \delta_n \mathbf{f}_n^T K^T) \mathbf{1}_n - \delta_n y_n) \\ &= \frac{1}{1 - \delta_n^2} (y_n + \mathbf{f}_n^T \mathbf{1}_n - \delta_n \mathbf{f}_n^T K \mathbf{1}_n - \delta_n y_n). \end{aligned}$$

Since from the definition of  $K$ ,  $K \mathbf{1} = \mathbf{1}$ , we get that

$$\begin{aligned} y_{n+1} + \mathbf{f}_{n+1}^T \mathbf{1}_{n+1} &= \frac{1}{(1 - \delta_n)(1 + \delta_n)} (1 - \delta_n) (y_n + \mathbf{f}_n^T \mathbf{1}_n) \\ &= \frac{1}{1 + \delta_n} (y_n + \mathbf{f}_n^T \mathbf{1}_n), \end{aligned}$$

which by (5.5) and the induction hypothesis implies that for  $X_0 \neq 0$  the first and last trades are nonzero and have the same sign as  $X_0$ .  $\square$

From Example 2 and 4, one might get the idea that for convex resilience functions, the optimal strategy is a convex function of time with a minimum in the middle of the trading interval. However, Example 5 shows that this is not true. The only thing we can prove is the following proposition.

**Proposition 8.** ([4]) *When  $G$  is convex and nonconstant, the optimal strategy  $\xi^*$  satisfies*

$$\begin{aligned} \xi_{t_0}^* &\geq \xi_{t_1}^* & \text{and} & & \xi_{t_{N-1}}^* &\leq \xi_{t_N}^* & \text{if} & & X_0 > 0, \\ \xi_{t_0}^* &\leq \xi_{t_1}^* & \text{and} & & \xi_{t_{N-1}}^* &\geq \xi_{t_N}^* & \text{if} & & X_0 < 0. \end{aligned}$$

*These inequalities are strict when  $G$  is strictly decreasing and the time grid is equidistant<sup>2</sup>.*

*Proof.* Since  $G$  is convex and nonconstant, by Proposition 3 it is positive definite, and hence by Theorem 1

$$\mathbf{x}^* = \frac{X_0}{\mathbf{1}^T M^{-1} \mathbf{1}} M^{-1} \mathbf{1} =: \lambda_0 M^{-1} \mathbf{1},$$

from which it follows that  $M\mathbf{x}^* = (\lambda_0, \dots, \lambda_0)^T$ . Equating the first two components of this equality, we get that

$$\sum_{j=0}^N x_j^* G(|t_j - t_0|) = \sum_{j=0}^N x_j^* G(|t_j - t_1|),$$

i.e.

$$x_0^* G(0) + x_1^* G(t_1 - t_0) + \sum_{j=2}^N x_j^* G(|t_j - t_0|) = x_0^* G(t_1 - t_0) + x_1^* G(0) + \sum_{j=2}^N x_j^* G(|t_j - t_1|).$$

It follows that

$$(x_0^* - x_1^*) G(0) = (x_0^* - x_1^*) G(t_1 - t_0) + \sum_{j=2}^N x_j^* (G(|t_j - t_1|) - G(|t_j - t_0|)).$$

If  $X_0 > 0$ , by Theorem 2,  $\mathbf{x}^* \geq \mathbf{0}$ , and hence

$$(x_0^* - x_1^*) G(0) \geq (x_0^* - x_1^*) G(t_1 - t_0), \quad (5.6)$$

using the fact that  $G$  is nonincreasing. From this we easily find that

$$(x_0^* - x_1^*) (G(0) - G(t_1 - t_0)) \geq 0,$$

and hence  $x_0^* \geq x_1^*$ , again using that  $G$  is nonincreasing. Equating the last two components in  $M\mathbf{x}^* = (\lambda_0, \dots, \lambda_0)^T$ , we find in a similar way that  $x_{N-1}^* \leq x_N^*$ . If  $X_0 < 0$ , by Theorem 2,  $\mathbf{x}^* \leq \mathbf{0}$ , and hence  $x_0^* \leq x_1^*$  and  $x_{N-1}^* \geq x_N^*$ .

If  $G$  is (strictly) decreasing,  $G(|t_j - t_1|) - G(|t_j - t_0|) > 0$  for  $j \in \{2, \dots, N\}$ . In order to get a strict inequality in (5.6), at least one of the elements of  $\{x_2^*, \dots, x_N^*\}$  needs to be strictly positive. If we assume that the time grid is equidistant, we get from Proposition 7 that  $x_N^* > 0$  if  $X_0 > 0$  and  $x_N^* < 0$  if  $X_0 < 0$ . This yields  $x_0^* > x_1^*$  if  $X_0 > 0$  and  $x_0^* < x_1^*$  if  $X_0 < 0$ . Similarly, Proposition 7 says that  $x_0^* > 0$  if  $X_0 > 0$  and  $x_0^* < 0$  if  $X_0 < 0$ , yielding  $x_{N-1}^* < x_N^*$  if  $X_0 > 0$  and  $x_{N-1}^* > x_N^*$  if  $X_0 < 0$ .  $\square$

---

<sup>2</sup>In [4], equidistance is not explicitly assumed. However, in the proof the authors use Proposition 7, in which equidistance in fact is an assumption.



## Chapter 6

# Risk aversion

So far, we have focused on minimizing the expected execution cost of a trading strategy. However, in practice one is often interested in minimizing the corresponding risk as well. In this chapter, we look at how this can be done.

### 6.1 A mean-variance functional

Orders placed at a later point in time are exposed to a larger volatility risk than earlier placed orders, and are thus less interesting for traders who do not want to take on too much risk. Instead of minimizing the expected execution costs, it is therefore common to *maximize* a mean-variance functional such as

$$MV_\gamma(\boldsymbol{\xi}) := \mathbb{E}[R(\boldsymbol{\xi})] - \frac{\gamma}{2} \text{Var}(R(\boldsymbol{\xi})), \quad (6.1)$$

where  $\gamma \geq 0$  is a risk aversion parameter and  $R(\boldsymbol{\xi})$  denotes the revenues from the strategy  $\boldsymbol{\xi}$ , i.e.

$$R(\boldsymbol{\xi}) = \frac{-1}{2G(0)} \sum_{n=0}^N (S_{t_n+}^2 - S_{t_n}^2). \quad (6.2)$$

These revenues are just the negative costs of a strategy, as is clear from the definition of the expected execution costs in Chapter 2 (cf. (2.3)):

$$\mathbb{E}[R(\boldsymbol{\xi})] = -C(\boldsymbol{\xi}) = -X_0 S_0 - \mathbb{E}[C(\boldsymbol{\xi})]. \quad (6.3)$$

According to [4], when maximizing  $MV_\gamma(\boldsymbol{\xi})$  over all adapted strategies  $\boldsymbol{\xi}$ , a time-inconsistency arises, in the sense that the Bellman principle of optimality fails. To avoid this, one can restrict oneself to using deterministic strategies. Alternatively, one can maximize the expected utility  $\mathbb{E}[u(R(\boldsymbol{\xi}))]$ . In that case, it is more natural to consider revenues than costs, which is why we switched signs earlier.

Based on the main result in [29], the following proposition describes the equivalence of the two approaches under some conditions on the unaffected stock price process  $S^0$  and the utility function  $u^1$ .

---

<sup>1</sup>For a proof of this proposition, check [29].

**Proposition 9.** ([4]) Suppose that  $G$  is positive definite and that the unaffected price process  $S^0$  is both a martingale and a Lévy process with a finite exponential moment, i.e.  $\mathbb{E} \left[ e^{aS_1^0} \right] < \infty$  for some  $a > 0$ . Then the following conditions are equivalent for any strategy  $\xi^*$ :

- (a)  $\xi^*$  maximizes the expected utility  $\mathbb{E} \left[ -e^{\gamma R(\xi)} \right]$  in the class of all strategies  $\xi$ .
- (b)  $\xi^*$  is deterministic and maximizes  $MV_\gamma(\xi)$  in the class of all deterministic strategies  $\xi$ .

Based on this result, we will consider the maximization of  $MV_\gamma(\xi)$  over the class of deterministic strategies in more detail. We have the following central result.

**Proposition 10.** ([4]) Suppose the martingale  $S^0$  is such that  $\varphi(t) := \text{Var}(S_t^0) < \infty$  for all  $t \geq 0$ . Then, for any deterministic strategy  $\xi$ ,

$$MV_\gamma(\xi) = -X_0 S_0^0 - \overline{C}(\xi),$$

where  $\overline{C}$  is the symmetric quadratic form  $\overline{C}(\mathbf{x}) = \mathbf{x}^T \overline{M} \mathbf{x}$  for the matrix

$$\overline{M}_{ij} = \frac{1}{2}G(|t_i - t_j|) + \frac{\gamma}{2}\varphi(t_i \wedge t_j), \quad 0 \leq i, j \leq N. \quad (6.4)$$

*Proof.* From (6.1), (6.2) and (6.3), we have that

$$\begin{aligned} MV_\gamma(\xi) &= -X_0 S_0 - \mathbb{E}[C(\xi)] - \frac{\gamma}{2} \text{Var} \left( \frac{-1}{2G(0)} \sum_{n=0}^N (S_{t_{n+}}^2 - S_{t_n}^2) \right) \\ &= -X_0 S_0 - \frac{1}{2} \xi^T M \xi - \frac{\gamma}{2} \text{Var} \left( \frac{-1}{2G(0)} \sum_{n=0}^N (S_{t_{n+}}^2 - S_{t_n}^2) \right), \end{aligned}$$

where the expected value has disappeared because  $\xi$  is deterministic. We are interested in the last term. In the proof of Proposition 1, we showed that

$$\frac{1}{2G(0)} \sum_{n=0}^N (S_{t_{n+}}^2 - S_{t_n}^2) = \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 + C(\xi).$$

Now since  $\xi$  is deterministic, we get that

$$\text{Var} \left( \frac{-1}{2G(0)} \sum_{n=0}^N (S_{t_{n+}}^2 - S_{t_n}^2) \right) = \text{Var} \left( - \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \right) = \text{Var} \left( \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \right),$$

which in turn equals

$$\text{Var} \left( \sum_{n=0}^N \xi_{t_n} S_{t_n}^0 \right) = \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} \text{Cov} \left( S_{t_i}^0, S_{t_j}^0 \right).$$

Using the law of total expectation (which holds since  $S_t^0$  has a finite expected value and

variance) and the fact that  $(S_t^0)_{t \geq 0}$  is a martingale, we derive the following:

$$\begin{aligned}
\text{Cov}(S_{t_i}^0, S_{t_j}^0) &= \mathbb{E}[S_{t_i}^0 S_{t_j}^0] - \mathbb{E}[S_{t_i}^0] \mathbb{E}[S_{t_j}^0] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(S_{t_i \vee t_j}^0 - S_{t_i \wedge t_j}^0\right) S_{t_i \wedge t_j}^0 + \left(S_{t_i \wedge t_j}^0\right)^2 \mid \mathcal{F}_{t_i \wedge t_j}\right]\right] - (S_0^0)^2 \\
&= \mathbb{E}\left[S_{t_i \wedge t_j}^0 \mathbb{E}\left[S_{t_i \vee t_j}^0 - S_{t_i \wedge t_j}^0\right] + \left(S_{t_i \wedge t_j}^0\right)^2\right] - (S_0^0)^2 \\
&= \mathbb{E}\left[S_{t_i \wedge t_j}^0 (S_0^0 - S_0^0)\right] + \text{Var}\left(S_{t_i \wedge t_j}^0\right) + \mathbb{E}\left[S_{t_i \wedge t_j}^0\right]^2 - (S_0^0)^2 \\
&= \text{Var}\left(S_{t_i \wedge t_j}^0\right) + (S_0^0)^2 - (S_0^0)^2 \\
&= \text{Var}\left(S_{t_i \wedge t_j}^0\right)
\end{aligned}$$

Hence

$$\begin{aligned}
MV_\gamma(\xi) &= -X_0 S_0 - \frac{1}{2} \xi^T M \xi - \frac{\gamma}{2} \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} \text{Var}\left(S_{t_i \wedge t_j}^0\right) \\
&= -X_0 S_0^0 - \frac{1}{2} \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} G(|t_i - t_j|) - \frac{\gamma}{2} \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} \varphi(t_i \wedge t_j) \\
&= -X_0 S_0^0 - \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} \left(\frac{1}{2} G(|t_i - t_j|) - \frac{\gamma}{2} \varphi(t_i \wedge t_j)\right) \\
&= -X_0 S_0^0 - \sum_{i,j=0}^N \xi_{t_i} \xi_{t_j} \overline{M}_{ij} \\
&= -X_0 S_0^0 - \xi^T \overline{M} \xi.
\end{aligned}$$

□

The matrix with components  $\varphi(t_i \wedge t_j)$  is a covariance matrix, and therefore positive semidefinite. Hence, if  $G$  is positive definite, then  $\overline{M}$  is positive definite, and by Theorem 1 there exists a unique optimal strategy  $\xi^*$  in the class of deterministic strategies, given by

$$\xi^* = \frac{X_0}{\mathbf{1}^T \overline{M}^{-1} \mathbf{1}} \overline{M}^{-1} \mathbf{1}.$$

Theorem 1 also excludes standard price manipulation. However, to exclude transaction-triggered price manipulation, it is not sufficient that  $G$  is convex and nonconstant, as the following example shows.

**Example 8.** Consider the power-law resilience function  $G(t) = (1+t)^{-0.4}$ , and the covariance function  $\varphi(t) = \sigma^2 t^{1/5}$  with volatility  $\sigma = 0.3$ . Then  $G$  is convex, but  $\varphi$  is not (in fact,  $\varphi$  is concave). Using (2.6) and (6.4), we can compute the mean-variance optimal strategy, which is visualized in Figure 6.1 for a particular situation. We see that the optimal strategy has a negative component, even though  $G$  is convex and  $X_0 > 0$ .  $\triangle$

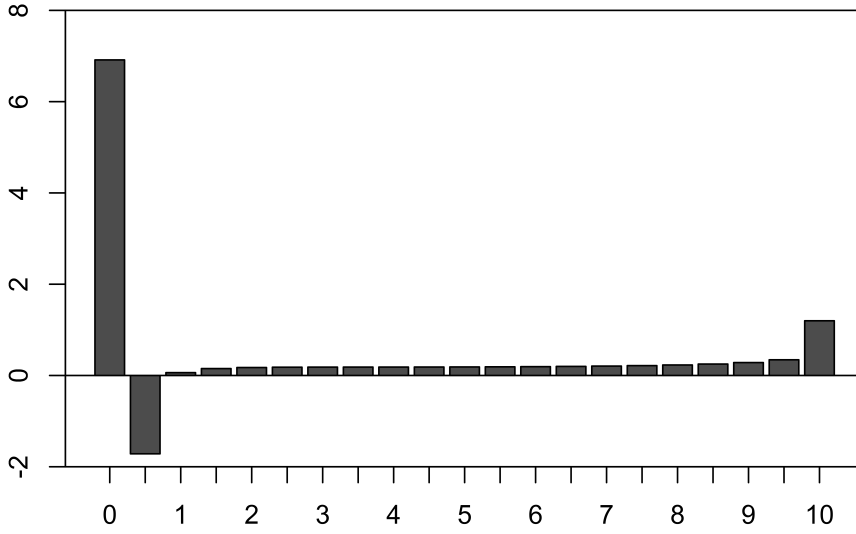


Figure 6.1: Mean-variance optimal strategy for power-law resilience  $G(t) = (1 + t)^{-0.4}$ , covariance function  $\varphi(t) = \sigma^2 t^{1/5}$  with  $\sigma = 0.3$ , risk aversion  $\gamma = 5$  and  $X_0 = 10$ ,  $T = 10$ ,  $N = 20$ .

## 6.2 Excluding transaction-triggered price manipulation

To exclude transaction-triggered price manipulation from the model, we need  $\varphi$  to be convex, as is stated the following theorem.

**Theorem 3.** ([4]) *Suppose that both  $G(t)$  and  $\varphi(t)$  are convex functions of  $t$ , and that  $G$  is nonconstant. Then there exists a unique strategy  $\xi^*$  maximizing the mean-variance functional  $MV_\gamma(\xi)$  in the class of all deterministic strategies  $\xi$ . Moreover,  $\xi^*$  has only nonnegative components for a buy program and nonpositive components for a sell program. All components of  $\xi^*$  are strictly positive (negative) as soon as  $G$  is, in addition, strictly convex.*

*Proof.* Since  $\varphi(t)$  is a convex function of  $t$ , it must be continuous except possibly in  $t = 0$ , where we know that  $\varphi(0) = \text{Var}(S_0^0) = 0$ . By Theorem 1.36 in [27], it follows that  $\varphi(t) = \text{Var}(S_t^0) < \infty$  on every closed and bounded interval  $[0, T]$ . Hence, by Proposition 10, maximizing the mean-variance functional  $MV_\gamma$  over deterministic strategies is equivalent to minimizing the quadratic form  $\bar{C}(\mathbf{x}) = \mathbf{x}^T \bar{M} \mathbf{x}$ , where  $\bar{M}$  is the matrix defined by

$$\bar{M}_{ij} = \frac{1}{2}G(|t_i - t_j|) + \frac{\gamma}{2}\varphi(t_i \wedge t_j), \quad 0 \leq i, j \leq N.$$

Being a covariance matrix, the matrix with components  $\varphi(t_i \wedge t_j)$  is positive semidefinite. Since  $G$  is convex and nonconstant by assumption, and hence positive definite by Proposition 3, the matrix  $\bar{M}$  is positive definite as well. Therefore, by Theorem 1, there exists a unique minimizer  $\xi^*$  of  $\bar{C}(\xi)$  under the constraint  $\xi^T \mathbf{1} = X_0$ . We proceed as in the proof of Theorem 2 to show that there exists no vector  $\mathbf{z}$  such that  $\mathbf{z}^T \mathbf{1} = 0$  and  $\bar{M} \mathbf{z} > \mathbf{0}$ . Then  $\bar{M}^{-1} \mathbf{1} \geq \mathbf{0}$  or  $\bar{M}^{-1} \mathbf{1} \leq \mathbf{0}$  and hence all components of  $\xi^*$  have the same sign.

To exclude the existence of a vector  $\mathbf{z}$  satisfying  $\mathbf{z}^T \mathbf{1} = 0$  and  $\overline{M}\mathbf{z} > \mathbf{0}$ , we first do the trivial observation

$$\varphi(t_N \wedge t_m) = \varphi(t_{N-1} \wedge t_m) \text{ for } m \in \{0, \dots, N-1\}.$$

Next, since  $\varphi(t) = \text{Var}(S_t^0) < \infty$ , the law of total variance gives that for  $t \geq s$ ,

$$\text{Var}(S_t^0) = \mathbb{E} \left[ \text{Var}(S_t^0 \mid \mathcal{F}_s) \right] + \text{Var} \left( \mathbb{E}[S_t^0 \mid \mathcal{F}_s] \right) = \mathbb{E} \left[ \text{Var}(S_t^0 \mid \mathcal{F}_s) \right] + \text{Var}(S_s^0),$$

where the last equality follows from the fact that  $(S_t^0)_{t \geq 0}$  is a martingale. We get that for  $t \geq s$ ,  $\text{Var}(S_t^0) \geq \text{Var}(S_s^0)$ , and hence  $\varphi(t) := \text{Var}(S_t^0)$  is nondecreasing. This implies that

$$\varphi(t_m \wedge t_0) \leq \varphi(t_m \wedge t_1) \text{ for } m \in \{1, \dots, N\}.$$

Finally, if  $t_k > t_l$  for  $1 \leq k \leq N-1$  and  $0 \leq l \leq N$ , then  $t_{k-1} \geq t_l$ , and clearly  $t_{k+1} > t_l$ . On the other hand, if  $t_k < t_l$ , then  $t_{k+1} \leq t_l$ , and clearly  $t_{k-1} < t_l$  (as we saw in the proof of Theorem 2). Hence, for  $l \neq k$ ,

$$\begin{aligned} \varphi(t_k \wedge t_l) &= \begin{cases} \varphi(t_l) = \alpha \varphi(t_l) + (1-\alpha) \varphi(t_l) & \text{if } t_k > t_l \\ \varphi(t_k) \leq \alpha \varphi(t_{k-1}) + (1-\alpha) \varphi(t_{k+1}) & \text{if } t_k < t_l \end{cases} \\ &= \begin{cases} \alpha \varphi(t_{k-1} \wedge t_l) + (1-\alpha) \varphi(t_{k+1} \wedge t_l) & \text{if } t_k > t_l \\ \alpha \varphi(t_{k-1} \wedge t_l) + (1-\alpha) \varphi(t_{k+1} \wedge t_l) & \text{if } t_k < t_l \end{cases} \end{aligned}$$

from which it follows directly that

$$\varphi(t_k \wedge t_l) \leq \alpha \varphi(t_{k-1} \wedge t_l) + (1-\alpha) \varphi(t_{k+1} \wedge t_l).$$

This means that  $\overline{M}_{ij} = \frac{1}{2}G(|t_i - t_j|) + \frac{\gamma}{2}\varphi(t_i \wedge t_j)$  satisfies (4.2), (4.4) and (4.6), and hence  $\overline{M}$  satisfies (4.3), (4.5) and (4.7). Therefore, we can repeat the proof of Theorem 2, replacing the matrix  $M$  by  $\overline{M}$ , and conclude that all components of  $\xi^*$  have the same sign. For strictly convex  $G$ , the proof of Theorem 2 gives us in addition that all components of  $\xi^*$  are nonzero.  $\square$

The following example gives a nice illustration of the result in Theorem 3.

**Example 9.** Consider again the convex power-law resilience function  $G(t) = (1+t)^{-0.4}$ , but now in combination with the covariance function  $\varphi(t) = \sigma^2 t$ , which is convex as well. The mean-variance optimal strategy in this case is shown in Figure 6.2 for a particular time grid and various values of the risk aversion parameters  $\gamma$ . We see how adding risk aversion to the model disturbs the time reversibility of the optimal strategy. Later trades become less interesting for the trader, which is due to the associated higher volatility risk.

$\triangle$

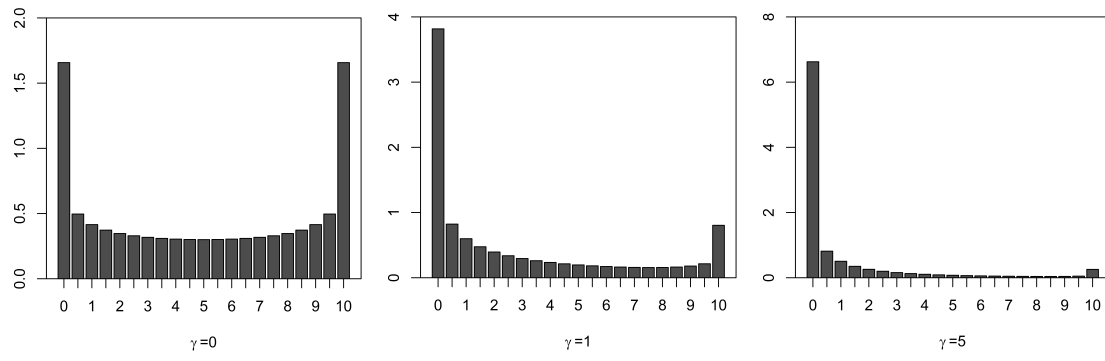


Figure 6.2: Mean-variance optimal strategy for power-law resilience  $G(t) = (1+t)^{-0.4}$ , covariance function  $\varphi(t) = \sigma^2 t$  with  $\sigma = 0.3$ , various risk aversion parameters  $\gamma$  and  $X_0 = 10$ ,  $T = 10$ ,  $N = 20$ .

## Chapter 7

# A transient linear model in continuous time

In the previous chapters we studied a discrete-time market impact model of the second generation. In Chapter 2, we excluded standard price manipulation and found an optimal trading strategy. In Chapter 4, we excluded transaction-triggered price manipulation as well. We also looked at some qualitative properties, and discussed the concept of risk aversion.

In this chapter, we will consider a transient linear price impact model in continuous time. In this model, trades can be made at all times, not only at specified points in time like in the discrete-time case. Many of the results we will discuss are similar to their discrete-time counterparts. However, we will see that in some cases we need extra assumptions, for instance in order to guarantee the existence of an optimal trading strategy.

### 7.1 Replicating the discrete second-generation model

Before introducing the model, we need to make some assumptions in order to assure that the unaffected stock price process is sufficiently well-behaved. We follow [18] in assuming that the unaffected price process  $(S_t^0)_{t \geq 0}$  is a right-continuous martingale defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  that satisfies the usual conditions. By this, it is meant that (cf. [21])

- $\mathcal{F}$  is  $\mathbb{P}$ -complete, i.e. if  $\mathcal{B} \subset \mathcal{A} \in \mathcal{F}$  and  $\mathbb{P}(\mathcal{A}) = 0$ , then  $\mathcal{B} \in \mathcal{F}$  and  $\mathbb{P}(\mathcal{B}) = 0$ ,
- $\mathcal{F}_0$  contains all  $\mathbb{P}$ -null sets,
- the filtration is right-continuous, i.e.  $\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$  for all  $t$ .

We also assume that  $\mathcal{F}_0$  is  $\mathbb{P}$ -trivial, meaning that any  $\mathcal{F}_0$ -measurable random variable is constant  $\mathbb{P}$ -a.s.

When we dealt with discrete-time strategies, we focused on trades  $\xi_{t_n}$ , describing the amount of shares that was bought or sold at time  $t_n$ . In continuous time, we will instead consider a stochastic process  $(X_t)_{t \geq 0}$ , representing the number of shares held by the trader at each time  $t$ . Along with [18], we assume that strategies are admissible in the following sense:

- the function  $t \rightarrow X_t$  is left-continuous and adapted,
- the function  $t \rightarrow X_t$  has finite and  $\mathbb{P}$ -a.s. bounded total variation,
- there exists  $T > 0$  such that  $X_t = 0$   $\mathbb{P}$ -a.s. for all  $t > T$ .

We make a few comments on how to interpret these conditions. The first condition is clear: we want  $X_t$  to depend on information available at time  $t$  only, and information about a trade first becomes available right after this trade is executed. Next, remember that trades usually are executed over a certain time horizon  $[0, T]$ , like we saw for the discrete-time case. Excluding the possibility of (transaction-triggered) price manipulation, a sell order thus corresponds to a nonincreasing asset position  $(X_t)_{t \geq 0}$  for which  $X_0 = x$  and  $X_{T+} = 0$ . (Since  $X_t$  is left-continuous,  $X_T$  can be nonzero even if  $X_{T+} = 0$ .) Likewise, a buy order corresponds to a nondecreasing asset position  $(Y_t)_{t \geq 0}$  with  $Y_0 = -y$  and  $Y_{T+} = 0$ . In other words, we interpret buying as filling a short position via a buy order, and we will continue to do so henceforth. This explains the third condition, which may seem a little strange at first sight.

A general trading strategy  $(Z_t)_{t \geq 0}$  may consist of alternating buy and sell orders. Such a strategy can be described as the sum of a nonincreasing strategy  $(X_t)_{t \geq 0}$  and a nondecreasing strategy  $(Y_t)_{t \geq 0}$ , i.e.  $Z_t = X_t + Y_t$ . But the existence of this representation is, by the definition of total variation, equivalent to requiring that  $(Z_t)_{t \geq 0}$  is of bounded total variation, hence the second condition.

Finally, the authors of [18] remark that by the third condition, the total variation of  $(Z_t)_{t \geq 0}$  is bounded by  $X_0 + |Y_0|$ . Since  $X_0 = x$  and  $|Y_0| = y$  are the total sizes of all sell and buy orders respectively, the second condition is equivalent to requiring that the total sizes of both buy and sell orders is bounded. From an economic point of view, this assumption can be made without loss of generality, as any stock only has a finite number of shares.

We have now laid a solid foundation to build our model on. Using the admissible strategy  $(X_t)_{t \geq 0}$ , the price impact model, as introduced in [18], is given by

$$S_t = S_t^0 + \int_{\{s < t\}} G(t-s) dX_s, \quad (7.1)$$

where the *decay kernel*<sup>1</sup>  $G : (0, \infty) \rightarrow [0, \infty)$  is a measurable function. For this expression to make sense, we assume that<sup>2</sup>

$$G \text{ is bounded and } G(0) := \lim_{t \downarrow 0} G(t) \text{ exists.} \quad (7.2)$$

Since  $(X_t)_{t \geq 0}$  is a stochastic process, the price process is given by a stochastic integral, which is well-defined since  $(X_t)_{t \geq 0}$  can be built up by a local martingale and an adapted process of bounded variation, making it a semimartingale. However, we assumed that  $X_t$  has finite and  $\mathbb{P}$ -a.s. bounded total variation as a function of  $t$ , which means that the local martingale part of  $(X_t)_{t \geq 0}$  vanishes. We are thus not really dealing with an Itô integral, but should rather say that we  $\mathbb{P}$ -a.s. end up with a (deterministic) Lebesgue-Stieltjes integral.

<sup>1</sup>This is the continuous-time version of what was called resilience function in the discrete-time model.

<sup>2</sup>This assumption can be relaxed so as to include weakly singular decay kernels like  $G(t) = t^{-\gamma}$  for  $0 < \gamma < 1$ . This relaxation is carried out in [18], but we will not look deeper into this.



Comparing the model (7.1) with its discrete-time counterpart

$$S_t = S_t^0 + \sum_{t_n < t} \xi_{t_n} G(t - t_n),$$

as given in (2.1), we see that they are logical equivalents. Note that  $dX_s := X_{s+} - X_s$  denotes the change in shares held by the trader in the infinitesimal time interval  $[s, s + \delta]$  for  $\delta > 0$ , i.e. the size of the trade executed at time  $s$ .

## 7.2 Expected execution costs

As for the discrete-time model, we will now try to define the expected execution cost of a trading strategy  $X$ . When  $X$  jumps at time  $t$ , we are in the discrete-time situation, and (2.2) applies. That is, the price is moved from  $S_t$  to  $S_{t+} = S_t + G(0) \Delta X_t$ , and the costs become

$$\begin{aligned} \int_{S_t}^{S_{t+}} \frac{S}{G(0)} dS &= \frac{1}{2 G(0)} (S_{t+}^2 - S_t^2) \\ &= \frac{1}{2 G(0)} \left( G(0)^2 (\Delta X_t)^2 + 2 G(0) S_t \Delta X_t \right) \\ &= \frac{G(0)}{2} (\Delta X_t)^2 + S_t \Delta X_t. \end{aligned}$$

Next, suppose  $X$  is continuous in  $t$ . Then, contrary to the discrete-time situation, an infinitesimal order  $dX_t$  may be executed at time  $t$ . The associated asset price is  $S_t$ , yielding an execution cost  $S_t dX_t$ . This means that the total cost of a continuous strategy is given by

$$\lim_{\Delta t \rightarrow 0} \sum_t S_t \Delta X_t = \int S_t dX_t = \int S_t^0 dX_t + \int \int_{\{s < t\}} G(t - s) dX_s dX_t. \quad (7.3)$$

Remark that using a continuous strategy is only a theoretical possibility, as in practice, one cannot buy or sell an infinitesimal amount of shares.

For an arbitrary strategy  $X$ , the total cost is given by the sum of the costs of the points where  $X$  jumps and where  $X$  is continuous. Let  $\mathcal{J}$  denote the collection of time points where  $X$  jumps and  $\mathcal{C}$  the times where  $X$  is continuous. Then, the total cost of an admissible strategy  $X$  is given by

$$\lim_{\Delta t \rightarrow 0} \sum_{t \in \mathcal{C}} S_t \Delta X_t + \sum_{t \in \mathcal{J}} S_t \Delta X_t + \sum_{t \in \mathcal{J}} \frac{G(0)}{2} (\Delta X_t)^2 = \int S_t dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2,$$

where the last sum should be understood as taken over all jumps, i.e. over all  $t \in \mathcal{J}$ .

The expected execution cost of any admissible strategy is now given by the following proposition, which can be regarded as the continuous-time version of Proposition 1.

**Proposition 11.** ([18]) *The expected execution costs of an admissible strategy  $X$  are*

$$\mathbb{E} \left[ \int S_t dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2 \right] = -S_0^0 X_0 + \frac{1}{2} \mathbb{E}[C(X)],$$

where

$$C(X) := \int \int G(|t - s|) dX_s dX_t.$$

*Sketch of proof.* By (7.3),

$$\int S_t \, dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2 = \int S_t^0 \, dX_t + \int \int_{\{s < t\}} G(t-s) \, dX_s \, dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2.$$

Similarly to the proof of Proposition 1, we observe that

$$\begin{aligned} \int \int_{\{s < t\}} G(t-s) \, dX_s \, dX_t &= \frac{1}{2} \int \int G(|t-s|) \, dX_s \, dX_t - \frac{1}{2} \int G(0) \, \Delta X_t \, dX_t \\ &= \frac{1}{2} \int \int G(|t-s|) \, dX_s \, dX_t - \frac{G(0)}{2} \sum (\Delta X_t)^2. \end{aligned}$$

Hence

$$\begin{aligned} \int S_t \, dX_t + \frac{G(0)}{2} \sum (\Delta X_t)^2 &= \int S_t^0 \, dX_t + \frac{1}{2} \int \int G(|t-s|) \, dX_s \, dX_t \\ &= \int S_t^0 \, dX_t + \frac{1}{2} C(X), \end{aligned}$$

and it remains to show that  $\mathbb{E} \left[ \int S_t^0 \, dX_t \right] = -S_0^0 X_0$ . Since  $X$  and  $S^0$  might jump at the same time, we cannot perform an integration by parts. Instead, we can use the fact that  $X$  is an optional martingale and get that

$$\mathbb{E} \left[ \int_{[0,T]} S_t^0 \, dX_t \right] = \mathbb{E} \left[ \int_{[0,T]} S_T^0 \, dX_t \right] = \mathbb{E} [S_T^0 (X_T - X_0)] = -S_0^0 X_0,$$

which is similar to (2.5). For the complete argument, check the proof of Lemma 2.3 in [18].  $\square$

As for the discrete-time model, we have to discuss the issue of viability. We mentioned in Chapter 2 that, in order to exclude arbitrage opportunities, our model should not admit price manipulation. Again, a price manipulation strategy is defined as a round trip with strictly negative expected execution cost. In the continuous-time case, a round trip is defined as an admissible strategy with  $X_0 = 0$ . Since we interpret buying as filling a short position, a round trip is a strategy for which the trader ends up with the same portfolio as he started with. Clearly, price manipulation is excluded as soon as  $C(X) \geq 0$  for all admissible strategies. We will continue to call  $G$  positive semidefinite when  $C(X) \geq 0$  for all admissible strategies  $X$ , and positive definite when  $C(X) > 0$  for every nonzero admissible strategy  $X$ . The following extension to Bochner's theorem characterizes all positive semidefinite decay kernels in the class of continuous and bounded functions.

**Proposition 12.** ([18]) *Let  $G$  be continuous and bounded. Then we have  $C(X) \geq 0$  for all admissible strategies  $X$  if and only if  $G(|\cdot|)$  can be represented as the Fourier transform of a positive finite Borel measure  $\mu$  on  $\mathbb{R}$ , i.e.,*

$$G(|x|) = \int e^{ixz} \, \mu(dz).$$

*If, in addition, the support of  $\mu$  is not discrete, then  $C(X) > 0$  for every nonzero admissible strategy  $X$ .*

*Sketch of proof.* Suppose first that  $C(X) \geq 0$  for all admissible strategies  $X$ . Then this holds in particular for all admissible strategies  $X$  with discrete support, which brings us in the context of Chapter 2. Bochner's theorem thus gives that  $G(| \cdot |)$  is the Fourier transform of a positive finite Borel measure  $\mu$  on  $\mathbb{R}$ .

Now suppose conversely that  $G(|x|) = \int e^{ixz} \mu(dz)$ . Then, for any admissible strategy  $X$ ,

$$C(X) = \int \int \int e^{iz(t-s)} \mu(dz) dX_s dX_t,$$

which by the Fubini-Tonelli theorem implies that

$$\begin{aligned} C(X) &= \int \left( \int e^{izt} dX_t \right) \left( \int e^{-izs} dX_s \right) \mu(dz) \\ &= \int \left( \int e^{izt} dX_t \right) \overline{\left( \int e^{izs} dX_s \right)} \mu(dz) \\ &= \int |\widehat{X}(z)|^2 \mu(dz) \geq 0, \end{aligned}$$

where  $\widehat{X}(z) = \int e^{itz} dX_t$  is the Fourier-Stieltjes transform of the strategy  $X$ , which is well-defined since  $dX$  has compact support.

To prove that  $C(X) > 0$  for every nonzero strategy  $X$  when the support of  $\mu$  is not discrete, one can show that the zero set of  $\widehat{X}$  must be discrete. This implies that the support of  $\mu$  needs to be discrete in order for  $C(X) = \int |\widehat{X}(z)|^2 \mu(dz)$  to equal zero. For the complete argument, see the proof of Proposition 2.6 in [18].  $\square$

In the sequel, to exclude price manipulation, we will assume that  $G$  is positive semidefinite. By Property 3 on page 243 in [19], assumption (7.2) is then automatically fulfilled. In the discrete-time case, positive semidefiniteness was sufficient for an optimal strategy to exist, but we will see that this is not the case for strategies in continuous time. What does stay the same, is the fact that every convex and nonconstant  $G$  is positive definite, as proved in Chapter 3.

### 7.3 A Fredholm integral equation

We will now consider the continuous-time optimal trade execution problem of minimizing the expected execution cost of an admissible strategy  $X$ . Following [18], we require that the support of our strategies is contained in a given compact (i.e. closed and bounded) set  $\mathbb{T} \subset [0, \infty)$ . When we choose a discrete  $\mathbb{T}$  with a fixed time horizon  $T$ , i.e.  $\mathbb{T} = [0, T]$ , we are in the context of Chapter 2 to 6.

In Chapter 2, we saw that in discrete time, the optimal strategy is deterministic. This is also true for the continuous-time case. By Proposition 11, every admissible strategy minimizing the expected execution costs must have sample paths in the class

$$\mathcal{X}(x, \mathbb{T}) := \left\{ X \mid X \text{ a deterministic admissible strategy with } X_0 = x \text{ and support in } \mathbb{T} \right\}.$$

An argument for this is given in [18]. It follows that we can restrict ourselves to studying the minimization of  $C(\cdot)$  over the strategies in  $\mathcal{X}(x, \mathbb{T})$ . Since only the market movements enter the cost function defined in Proposition 11, and not the market situation itself, this

does not seem unreasonable. As in the discrete-time case, deterministic minimizers of  $C(\cdot)$  will be called optimal strategies.

The authors of [18] mention that for applications, such as implementation in industrial order execution algorithms, one might prefer strategies that are adapted to the stock price process. They claim that such adapted strategies arise when, instead of minimizing the expected execution costs, one considers maximizing the expected utility of the revenues as mentioned in Chapter 6. However, their (and our) goal is to verify the regularity of the model (7.1), and characterizing those decay kernels for which well-behaved strategies exist.

We now do an important observation: any optimal strategy, if existing, is unique as long as  $G$  is positive definite.

**Theorem 4.** ([18]) *When  $G$  is positive definite, there exists at most one optimal strategy for given  $x$  and  $\mathbb{T}$ .*

*Proof.* We start with a definition that we will use throughout this and the following chapter: for admissible strategies  $X$  and  $Y$ , let

$$C(X, Y) := \int \int G(|t - s|) dX_s dY_t.$$

This definition gives rise to the characterization  $C(X) = C(X, X)$ . By the Fubini-Tonelli theorem (cf. Theorem 2.37 in [15]),  $C(X, Y) = C(Y, X)$ , so we have symmetry. Furthermore, for admissible  $Z$  (note that the sum of two admissible strategies is again admissible),

$$\begin{aligned} C(X + Z, Y) &= \int \int G(|t - s|) d(X + Z)_s dY_t \\ &= \int \int G(|t - s|) dX_s dY_t + \int \int G(|t - s|) dZ_s dY_t \\ &= C(X, Y) + C(Z, Y), \\ C(X, Y + Z) &= C(X, Y) + C(X, Z), \\ C(\lambda X, Y) &= \lambda C(X, Y) \\ &= C(X, \lambda Y), \end{aligned}$$

and it follows that  $C(X, Y)$  is a symmetric bilinear form. We also have the polarization identity

$$C(X, Y) = \frac{1}{2} \left( C(X + Y) - C(X) - C(Y) \right) \quad (7.4)$$

as a result of the following computation:

$$\begin{aligned} C(X + Y) &= C(X + Y, X + Y) \\ &= C(X, X + Y) + C(Y, X + Y) \\ &= C(X, X) + C(X, Y) + C(Y, X) + C(Y, Y) \\ &= C(X) + 2 C(X, Y) + C(Y). \end{aligned}$$

This polarization identity we will use to show that  $X \mapsto C(X)$  is strictly convex, which implies the uniqueness of optimal execution strategies when they exist. First note that

$$\begin{aligned} C(X - Y) &= 2 C(X, -Y) + C(X) + C(-Y) \\ &= C(X) + C(-Y, -Y) - 2 C(X, Y) \\ &= C(X) + C(Y) - 2 C(X, Y). \end{aligned}$$

Now for  $X \neq Y$ ,  $X - Y$  is a nonzero strategy, hence  $C(X - Y) > 0$  by positive definiteness of  $G$ . This yields

$$2 C(X, Y) < C(X) + C(Y). \quad (7.5)$$

We find that

$$\begin{aligned} C\left(\frac{1}{2}X + \frac{1}{2}Y\right) &= 2 C\left(\frac{1}{2}X, \frac{1}{2}Y\right) + C\left(\frac{1}{2}X\right) + C\left(\frac{1}{2}Y\right) \\ &= \frac{1}{2} C(X, Y) + C\left(\frac{1}{2}X, \frac{1}{2}X\right) + C\left(\frac{1}{2}Y, \frac{1}{2}Y\right) \\ &= \frac{1}{4} \left( 2 C(X, Y) + C(X) + C(Y) \right), \end{aligned}$$

which by (7.5) implies that

$$C\left(\frac{1}{2}X + \frac{1}{2}Y\right) < \frac{1}{2} C(X) + \frac{1}{2} C(Y).$$

This shows that  $X \mapsto C(X)$  is strictly mid-point convex. However, for noncontinuous functions like

$$f(x) = \begin{cases} x^2 & \text{for } x \in \mathbb{Q} \\ x^2 + 1 & \text{for } x \notin \mathbb{Q}, \end{cases}$$

this does not imply convexity, so we have to show continuity. But this follows directly from the fact that

$$\alpha \mapsto C(\alpha X + (1 - \alpha)Y) = \alpha^2 C(X) + 2\alpha(1 - \alpha) C(X, Y) + (1 - \alpha)^2 C(Y)$$

is continuous. □

In Proposition 6, we used the uniqueness of the optimal strategy in discrete time to prove time reversibility. We can do something similar in continuous time.

**Corollary 2. (Time reversibility, [18])** *When  $G$  is strictly positive definite and  $\mathbb{T}$  is reversible, then the optimal strategy is time reversible.*

*Proof.* Let  $T := \max \mathbb{T}$  and suppose for simplicity that  $\min \mathbb{T} = 0$ . Then the reversal  $\check{\mathbb{T}}$  of  $\mathbb{T}$  is defined as  $\check{\mathbb{T}} := \{T - t \mid t \in \mathbb{T}\}$ . When  $\check{\mathbb{T}} = \mathbb{T}$ , we call  $\mathbb{T}$  reversible. An example of a reversible time set is the interval  $[0, T]$ . For a strategy  $X \in \mathcal{X}(x, \mathbb{T})$ , the time reversal is defined as

$$\check{X}_t := \begin{cases} x - X_{(T-t)+} & \text{for } t \leq T, \\ 0 & \text{for } t > T. \end{cases}$$

Remark how this definition preserves the left-continuity. Clearly,  $\check{X}_t \in \mathcal{X}(x, \check{\mathbb{T}})$  and

$$\begin{aligned} C(\check{X}) &= \int \int G(|t-s|) \, d\check{X}_s \, d\check{X}_t \\ &= \int \int G(|t-s|) \, d(x - X_{(T-s)+}) \, d(x - X_{(T-t)+}) \\ &= \int \int G(|t-s|) \, dX_{(T-s)+} \, dX_{(T-t)+} \\ &= \int \int G(|t-s|) \, dX_s \, dX_t = C(X). \end{aligned}$$

This implies that  $\check{X}^*$  is optimal in  $\mathcal{X}(x, \check{\mathbb{T}})$  if and only if  $X^*$  is optimal in  $\mathcal{X}(x, \mathbb{T})$ . For reversible  $\mathbb{T}$ , the time-reversed optimal strategy  $\check{X}^*$  is thus again optimal in  $\mathcal{X}(x, \mathbb{T})$ . Therefore, when  $G$  is strictly positive definite, Theorem 4 gives that  $\check{X}^* = X^*$ .  $\square$

With uniqueness guaranteed for positive definite decay kernels, we continue to search for optimal strategies. Characterizing optimal strategies as measure-valued solutions of a generalized Fredholm integral equation of the first kind, the following theorem is the main result of this chapter.

**Theorem 5.** ([18]) *Suppose that  $G$  is positive semidefinite. Then  $X^* \in \mathcal{X}(x, \mathbb{T})$  minimizes  $C(\cdot)$  over  $\mathcal{X}(x, \mathbb{T})$  if and only if there is a constant  $\lambda$  such that  $X^*$  solves the generalized Fredholm integral equation*

$$\int G(|t-s|) \, dX_s^* = \lambda \quad \text{for all } t \in \mathbb{T}. \quad (7.6)$$

*In this case<sup>3</sup>,  $C(X^*) = -\lambda x$ . In particular,  $\lambda$  must be nonzero as soon as  $G$  is strictly positive definite and  $x \neq 0$ .*

*Proof.* We start by proving that (7.6) is necessary for optimality. To this end, fix  $t_0, t \in \mathbb{T}$  and let  $X^*$  denote an optimal order execution strategy. Furthermore, let  $Y$  be the admissible round trip defined by

$$dY_s = \delta_{t_0}(ds) - \delta_t(ds),$$

i.e. the strategy consisting of buying one share at time  $t_0$  and selling one at time  $t$ . Then, since  $X^*$  is an optimal strategy, we get by (7.4) that for all  $\alpha \in \mathbb{R}$ ,

$$C(X^*) \leq C(X^* + \alpha Y) = C(X^*) + \alpha^2 C(Y) + 2\alpha C(X^*, Y).$$

As a function of  $\alpha$ , this entity is differentiable on  $\mathbb{R}$ . Now by optimality of  $X^*$ , we know that  $C(X^* + \alpha Y)$  attains a minimum for  $\alpha = 0$ , and hence

$$\left. \frac{d}{d\alpha} \right|_{\alpha=0} C(X^* + \alpha Y) = 2 C(X^*, Y) = 0$$

must hold. This yields

$$\begin{aligned} 0 &= C(X^*, Y) = C(Y, X^*) = \int \int G(|r-s|) \, dY_r \, dX_s^* \\ &= \int G(|t_0-s|) \, dX_s^* - \int G(|t-s|) \, dX_s^*, \end{aligned}$$

---

<sup>3</sup>In [18], it says that  $C(X^*) = \lambda x$ , but that has to be a misprint.

from which we get that

$$\int G(|t-s|) dX_s^* = \int G(|t_0-s|) dX_s^* =: \lambda.$$

By varying  $t \in \mathbb{T}$ , we see that the result holds for all  $t \in \mathbb{T}$ , which is what we had to prove.

It remains to show that (7.6) is sufficient for optimality. Suppose therefore that (7.6) holds for some  $X^* \in \mathcal{X}(x, \mathbb{T})$ , and let  $\tilde{X}$  be any other strategy in  $\mathcal{X}(x, \mathbb{T})$ . When we define  $Z := \tilde{X} - X^*$  and  $T := \max \mathbb{T}$ , we get that

$$Z_{T+} - Z_0 = \tilde{X}_{T+} - \tilde{X}_0 - (X_{T+}^* - X_0^*) = 0,$$

and hence

$$C(X^*, Z) = \int \int G(|t-s|) dX_s^* dZ_t = \int \lambda dZ_t = \lambda (Z_{T+} - Z_0) = 0.$$

By (7.4), it follows that

$$C(\tilde{X}) = C(X^* + Z) = C(X^*) + C(Z) + 2 C(X^*, Z) = C(X^*) + C(Z) \geq C(X^*),$$

which means that  $X^*$  is an optimal execution strategy. Furthermore,

$$C(X^*) = \int \int G(|t-s|) dX_s^* dX_t^* = \int \lambda dX_t^* = \lambda (X_{T+}^* - X_0^*) = -\lambda x. \quad \square$$

Theorem 5 allows us to find explicit optimal strategies for several choices of the decay kernel. It says that for positive semidefinite  $G$ , any strategy  $X^* \in \mathcal{X}(x, \mathbb{T})$  that satisfies (7.6) is optimal. We look at some examples.

**Example 10. (Permanent price impact)** As mentioned in Chapter 3, the constant decay kernel  $G \equiv 1$  is positive semidefinite as the Fourier transform of the Dirac measure  $\delta_0$ . Clearly, every strategy  $X \in \mathcal{X}(x, \mathbb{T})$  satisfies

$$\int G(|t-s|) dX_s = \int dX_s = X_{T+} - X_0 = -x \quad \text{for all } t.$$

Therefore, again all strategies are optimal and the costs are given by  $C(X) = x^2$ .  $\triangle$

**Example 11. (Exponential decay)** The exponential decay kernel  $G(t) = e^{-\rho t}$  for  $\rho > 0$ , which we studied in Chapter 3 as well, is convex and hence positive definite. The unique optimal execution strategy in  $\mathcal{X}(x, [0, T])$  is given as the solution of an optimal control problem (cf. Proposition 3 in [25]). Its dynamics are given by

$$dX_s^* = \frac{-x}{\rho T + 2} (\delta_0(ds) + \rho ds + \delta_T(ds)).$$

Indeed,

$$\begin{aligned} X_{T+}^* - X_0^* &= \int dX_s^* = \frac{-x}{\rho T + 2} \left( \int_{[0, T]} \delta_0(ds) + \int_{[0, T]} \rho ds + \int_0^T \delta_T(ds) \right) \\ &= \frac{-x}{\rho T + 2} (1 + \rho T + 1) = -x, \end{aligned}$$

and  $X^* \in \mathcal{X}(x, [0, T])$ . Let us verify whether  $X^*$  solves the Fredholm integral equation (7.6). For all  $t$ , we have that

$$\begin{aligned}
\int G(|t-s|) dX_s^* &= \int_{[0,T]} e^{-\rho|t-s|} dX_s^* \\
&= \frac{-x}{\rho T + 2} \left( e^{-\rho t} + \int_0^t \rho e^{-\rho|t-s|} ds + e^{-\rho|t-T|} \right) \\
&= \frac{-x}{\rho T + 2} \left( e^{-\rho t} + \int_0^t \rho e^{-\rho(t-s)} ds + \int_t^T \rho e^{-\rho(s-t)} ds + e^{-\rho(T-t)} \right) \\
&= \frac{-x}{\rho T + 2} \left( e^{-\rho t} + e^{-\rho t} \int_0^t \rho e^{\rho s} ds + e^{\rho t} \int_t^T \rho e^{-\rho s} ds + e^{-\rho(T-t)} \right) \\
&= \frac{-x}{\rho T + 2} \left( e^{-\rho t} + e^{-\rho t} [e^{\rho s}]_{s=0}^t + e^{\rho t} [-e^{-\rho s}]_{s=t}^T + e^{-\rho(T-t)} \right) \\
&= \frac{-x}{\rho T + 2} \left( e^{-\rho t} + e^{-\rho t} (e^{\rho t} - 1) + e^{\rho t} (e^{-\rho t} - e^{-\rho T}) + e^{-\rho(T-t)} \right) \\
&= \frac{-2x}{\rho T + 2} =: \lambda.
\end{aligned}$$

So indeed,  $X^*$  satisfies (7.6), and  $\lambda \neq 0$  as soon as  $x \neq 0$ . Using that for  $0 < t \leq T$

$$X_{T+}^* - X_t^* = \int_{[t,T]} dX_s^* = \frac{-x}{\rho T + 2} (0 + \rho(T-t) + 1),$$

we obtain

$$X_t^* = x \frac{\rho(T-t) + 1}{\rho T + 2} \quad \text{for } 0 < t \leq T,$$

by putting  $X_{T+}^* = 0$ . This we can use to visualize the optimal strategy, as is done in Figure 7.1. Comparing with Figure 3.1, we see a similar pattern with so-called impulse trades at the beginning and end, and mutually identical trades in between.  $\triangle$



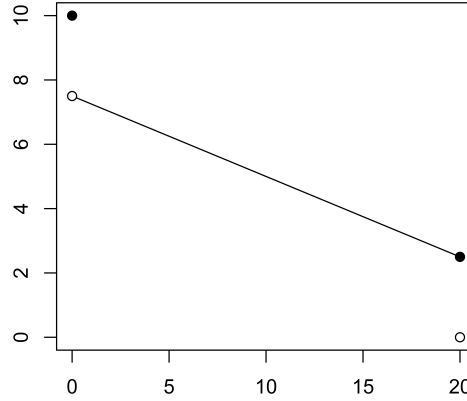


Figure 7.1: Optimal order execution strategy for the exponential decay function  $G(t) = e^{-t}$  on  $\mathbb{T} = [0, T]$  with  $T = 20$  and  $x = 10$ .

**Example 12. (Capped linear decay)** Consider the capped linear decay kernel  $G(t) = (1 - \rho t)^+$ , where  $\rho > 0$  is a positive constant. For  $\mathbb{T} = [0, T]$ , the optimal trading strategy is given by Proposition 13 below. In Figure 7.2, it is shown for specific values of  $\rho$  and  $T$ . From both Proposition 13 and Figure 7.2 we see that the optimal strategy is purely discrete.

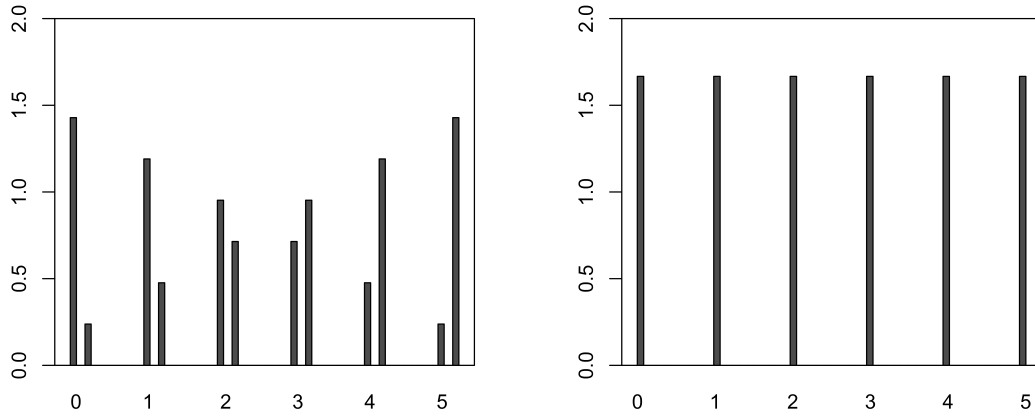


Figure 7.2: Optimal order execution strategies for the capped linear decay function  $G(t) = (1 - t)^+$  on  $\mathbb{T} = [0, T]$  with  $T = 5.2$  (left) and  $T = 5$  (right).

As in the discrete-time variant (cf. Figure 3.7), the optimal strategy is to wait until the price impact of one child order has vanished before executing the next. Now, in continuous time, the restriction that  $\rho = k/T$  for some  $k$  dividing  $N$  (cf. Proposition 4) can be dropped, resulting in two sequences of trades: one sequence starting from zero with the succeeding child orders executed immediately after the price impact of the previous

one has vanished, and one sequence ending at  $T$  with the preceding trades such that their effect has vanished by the time the next one is executed. If we take  $T$  and  $\rho$  such that for every  $i \in \{0, \dots, N\}$  there is a  $j \in \{0, \dots, N\}$  such that  $\delta_i/\rho = \delta_{T-j/\rho}$  (as in done in the right-hand side of Figure 7.2), then the paired trades coincide and we find the same optimal execution strategy as in Example 5 and Proposition 4.  $\triangle$

**Proposition 13.** ([18]) *The unique optimal strategy  $X^*$  for the capped linear decay kernel  $G(t) = (1 - \rho t)^+$  and  $\mathbb{T} = [0, T]$  corresponds to the purely discrete measure*

$$dX^* = \frac{-x}{N+2} \sum_{i=0}^N \left(1 - \frac{i}{N+1}\right) \left(\delta_{\frac{i}{\rho}} + \delta_{T-\frac{i}{\rho}}\right),$$

where  $N := \lfloor \rho T \rfloor$ .

*Proof.* <sup>4</sup> By Theorem 5, we need to show that

$$\int G(|t-s|) dX_s^*$$

is a constant function of  $t$ . This is equivalent to requiring that

$$\begin{aligned} g(t) &:= \frac{N+2}{-x} \int G(|t-s|) dX_s^* \\ &= \sum_{i=0}^N \left(1 - \frac{i}{N+1}\right) \left( G\left(\left|t - \frac{i}{\rho}\right|\right) + G\left(\left|T - t - \frac{i}{\rho}\right|\right) \right) \\ &= \sum_{i=0}^N \left(1 - \frac{i}{N+1}\right) \left( (1 - |\rho t - i|)^+ + (1 - |\rho(T-t) - i|)^+ \right) \end{aligned} \quad (7.7)$$

is a constant function of  $t$ . As  $g$  is continuous, it suffices to show that  $g$  is piecewise constant. Therefore, we divide the interval  $[0, T]$  in pieces, which we consider one by one. (Recall that we did something similar in Proposition 4 for the discrete-time version of the problem.) We define  $0 \leq \Delta < 1$  by  $\Delta := \rho T - N$ , and partition the interval  $[0, \rho T]$  in the following way:

$$0 \leq \Delta < 1 \leq 1 + \Delta < 2 \leq \dots < N \leq N + \Delta = \rho T.$$

*First case:*  $0 \leq \rho t \leq \Delta$ . In this case,  $0 \leq \rho t \leq 1$  and  $N \leq \rho(T-t) \leq N + \Delta$ . When  $|\rho t - i| \geq 1$  or  $|\rho(T-t) - i| \geq 1$  respectively, no contributions are made to the left-hand or right-hand side of the sum in (7.7). This means that in our case, only the terms for which  $i = 0, 1$  contribute to the left-hand side of this sum, and only the term for which  $i = N$  contributes to the right-hand side. We thus get

$$\begin{aligned} g(t) &= (1 - \rho t) + \left(1 - \frac{1}{N+1}\right) (1 - |\rho t - 1|)^+ + \left(1 - \frac{N}{N+1}\right) (1 - |\rho(T-t) - N|)^+ \\ &= 1 - \rho t + \left(1 - \frac{1}{N+1}\right) \rho t + \left(1 - \frac{N}{N+1}\right) (1 - \rho(T-t) + N). \end{aligned}$$

---

<sup>4</sup>In [18], there are some small mistakes in the proof of this proposition, but they do not affect the outcome.

When we only consider the terms containing  $\rho t$ , we are left with

$$\left(-1 + 1 - \frac{1}{N+1} + 1 - \frac{N}{N+1}\right)\rho t = 0.$$

This shows that  $g$  is constant for  $0 \leq \rho t \leq \Delta$ .

*Second case:*  $\Delta \leq \rho t \leq 1$ . In this case,  $N - 1 + \Delta \leq \rho(T - t) \leq N$  and again  $0 \leq \rho t \leq 1$ . Hence only the terms for which  $i \in \{0, 1, N - 1, N\}$  contribute to the sum in (7.7), yielding

$$\begin{aligned} g(t) = 1 - \rho t + \left(1 - \frac{1}{N+1}\right)\rho t + \left(1 - \frac{N-1}{N+1}\right)(1 - \rho(T-t) - (N-1)) \\ + \left(1 - \frac{N}{N+1}\right)(1 + \rho(T-t) - N). \end{aligned}$$

Adding again the coefficients of all terms containing  $\rho t$ , we get

$$-1 + 1 - \frac{1}{N+1} + 1 - \frac{N-1}{N+1} - 1 + \frac{N}{N+1} = 0,$$

and  $g$  is constant for  $\Delta \leq \rho t \leq 1$  as well.

*Third case:*  $k \leq \rho t \leq k + \Delta$  for  $k \in \{1, \dots, N-1\}$ . In this case, we have that  $k \leq \rho t \leq k+1$  and  $N - k \leq \rho(T-t) \leq N - k + \Delta$ , and only the terms corresponding to  $i \in \{k, k+1, N-k, N-k+1\}$  contribute to the sum in (7.7). We get

$$\begin{aligned} g(t) = \left(1 - \frac{k}{N+1}\right)(1 - \rho t + k) + \left(1 - \frac{k+1}{N+1}\right)(\rho t - k) \\ + \left(1 - \frac{N-k}{N+1}\right)(1 - \rho(T-t) + N - k) + \left(1 - \frac{N-k+1}{N+1}\right)(\rho(T-t) - N + k), \end{aligned}$$

and the sum of the coefficients of  $\rho t$  becomes again

$$-1 + \frac{k}{N+1} + 1 - \frac{k+1}{N+1} + 1 - \frac{N-k}{N+1} - 1 + \frac{N-k+1}{N+1} = 0.$$

*Fourth case:*  $k + \Delta \leq \rho t \leq k + 1$  for  $k \in \{1, \dots, N-1\}$ . In this case, we have that  $N - k - 1 + \Delta \leq \rho(T-t) \leq N - k$  and again  $k \leq \rho t \leq k+1$ . The only terms in (7.7) that do not cancel are those corresponding to  $i \in \{k, k+1, N-k-1, N-k\}$ . We thus get

$$\begin{aligned} g(t) = \left(1 - \frac{k}{N+1}\right)(1 - \rho t + k) + \left(1 - \frac{k+1}{N+1}\right)(\rho t - k) \\ + \left(1 - \frac{N-k-1}{N+1}\right)(N - k - \rho(T-t)) + \left(1 - \frac{N-k}{N+1}\right)(1 - N + k + \rho(T-t)), \end{aligned}$$

and adding the coefficients of the terms containing  $\rho t$  yields

$$-1 + \frac{k}{N+1} + 1 - \frac{k+1}{N+1} + 1 - \frac{N-k-1}{N+1} - 1 + \frac{N-k}{N+1} = 0.$$

*Fifth case:*  $N \leq \rho t \leq N + \Delta$ . This corresponds to  $0 \leq \rho(T-t) \leq \Delta$ . Since  $g(t) = g(T-t)$ , the result follows from the first case.

These five cases together show that  $g$  is piecewise constant on  $[0, T]$ , from which optimality follows. Uniqueness is guaranteed by Proposition 3 and Theorem 4.  $\square$

## 7.4 Nonexistence of optimal solutions

In the discrete-time case, positive semidefiniteness of  $G$  guaranteed the existence of an optimal trading strategy. The following theorem gives conditions for the *nonexistence* of an optimal solution in continuous time.

**Theorem 6.** ([18]) *Suppose that  $G(|\cdot|)$  can be represented as the Fourier transform of a positive finite Borel measure  $\mu$  for which*

$$\int e^{\varepsilon x} \mu(dx) < \infty \quad \text{for some } \varepsilon > 0. \quad (7.8)$$

*Suppose furthermore that the support of  $\mu$  is not discrete. Then there are no optimal strategies in  $\mathcal{X}(x, \mathbb{T})$  when  $x \neq 0$  and  $\mathbb{T}$  is not discrete.*

*Proof.* See the proof of Theorem 2.15 in [18]. □

In the following example we study two kernels which show that the result in Theorem 6 is closely related to the appearance of transaction-triggered price manipulation in discrete time, as studied in Chapter 3 and 4.

**Example 13.** The first examples we saw of transaction-triggered price manipulation, were the Gaussian and alternative power-law resilience functions (cf. Example 6 and 7). We noticed that the more refined the time grid was, the stronger the optimal strategies oscillated. From Theorem 6 it follows that in the most refined case, i.e. in continuous time, no optimal strategies exist for these kernels when  $x \neq 0$ .

Let us verify this. For the Gaussian decay kernel  $G(t) = \eta e^{-t^2}$ , we found that

$$G(t) = \int e^{itx} \mu(dx), \quad \text{where} \quad \mu(dx) = \frac{\eta}{\sqrt{4\pi}} e^{-x^2/4} dx.$$

This  $\mu$  clearly is a positive finite Borel measure, and

$$\begin{aligned} \int e^{\varepsilon x} \mu(dx) &= \int \frac{\eta}{\sqrt{4\pi}} e^{\varepsilon^2} e^{-(x-2\varepsilon)^2/4} dx \\ &= \eta e^{\varepsilon^2} \int \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-(x-2\varepsilon)^2/4} dx \\ &= \eta e^{\varepsilon^2} < \infty, \end{aligned}$$

where the last equality follows since the integral is the distribution function of the Gaussian distribution with standard deviation  $\sqrt{2}$ . As the support of  $\mu$  is not discrete, indeed, no optimal strategies exist when  $\mathbb{T}$  is not discrete.

The alternative power-law decay kernel  $G(t) = 1/(1+t^2)$  is the Fourier transform of the positive finite Borel measure  $\mu(dx) = \frac{1}{2}e^{-|x|}dx$ , as shown in Example 6. This gives us that

$$\begin{aligned} \int e^{\varepsilon x} \mu(dx) &= \int_{-\infty}^{\infty} \frac{1}{2} e^{\varepsilon x - |x|} dx \\ &= \frac{1}{2} \int_{-\infty}^0 e^{(\varepsilon+1)x} dx + \frac{1}{2} \int_0^{\infty} e^{(\varepsilon-1)x} dx \\ &= \frac{1}{2} \left[ \frac{1}{\varepsilon+1} e^{(\varepsilon+1)x} \right]_{x=-\infty}^0 + \frac{1}{2} \left[ \frac{1}{\varepsilon-1} e^{(\varepsilon-1)x} \right]_{x=0}^{\infty}. \end{aligned}$$

Choosing  $0 < \varepsilon < 1$ , this equals

$$\int e^{\varepsilon x} \mu(dx) = \frac{1}{2} \left( \frac{1}{\varepsilon + 1} - \frac{1}{\varepsilon - 1} \right) = \frac{1}{(1 - \varepsilon)(1 + \varepsilon)} < \infty.$$

Since  $\mu$  has a nondiscrete support in this case as well, no optimal strategies exist when  $\mathbb{T}$  is not discrete.  $\triangle$

As a variant of Theorem 6, we have the following corollary to the proof of that proposition. It also applies when the measure in (7.8) has discrete support.

**Corollary 3.** ([18]) *Suppose that  $G$  is the Fourier transform of a measure  $\mu$  for which condition (7.8) holds. Suppose furthermore that  $\mu(\{0\}) = 0$  and that  $\mathbb{T}$  is not discrete. Then every optimal strategy  $X^*$  in  $\mathcal{X}(x, \mathbb{T})$  satisfies  $C(X^*) = 0$ .*

*Proof.* This follows from the proof of Theorem 6. See the proof of Corollary 2.17 in [18].  $\square$

As an application of Corollary 3, we study the following decay kernel:

**Example 14. (Trigonometric decay)** The decay kernel

$$G(t) = \cos(\rho t)$$

is decreasing and nonnegative on  $[0, \frac{\pi}{2\rho}]$ . If we take  $0 < \rho \leq \frac{\pi}{2T}$ , we thus get a decreasing and nonnegative function on  $[0, T]$ . As

$$\begin{aligned} \int \frac{1}{2} e^{ity} (\delta_{-\rho} + \delta_{\rho}) &= \frac{1}{2} (e^{-i\rho t} + e^{i\rho t}) \\ &= \frac{1}{2} (\cos(\rho t) - i \sin(\rho t) + \cos(\rho t) + i \sin(\rho t)) \\ &= \cos(\rho t), \end{aligned}$$

$G(|t|) = G(t)$  is the Fourier transform of the discretely supported positive finite measure  $\mu = \frac{1}{2}(\delta_{-\rho} + \delta_{\rho})$ . For any  $\varepsilon > 0$ ,

$$\int e^{\varepsilon x} \mu(dx) = \int \frac{1}{2} e^{\varepsilon x} (\delta_{-\rho} + \delta_{\rho}) = \frac{1}{2} (e^{-\varepsilon\rho} + e^{\varepsilon\rho}) < \infty,$$

so clearly (7.8) is satisfied. Since also  $\mu(\{0\}) = 0$ , Corollary 3 applies when we for instance take  $\mathbb{T} = [0, T]$ . (Remark how Example 10 shows that the condition  $\mu(\{0\}) = 0$  is essential.) From Theorem 5, it follows that all optimal strategies must satisfy

$$\int_{[0, T]} \cos(\rho(t - s)) dX_s = 0 \quad \text{for } t \in [0, T]. \quad (7.9)$$

Along with [18], we now suppose for simplicity that  $\rho = 1$  and  $T = \frac{\pi}{2}$ , and consider the strategy  $X^1$  given by

$$dX_s^1 = -\frac{1}{2}\delta_0(ds) + \sin(s) ds - \frac{\pi}{4}\delta_{\pi/2}(ds).$$

This strategy is visualized in Figure 7.3. We see that the optimal strategy is not monotone, which means that there exists transaction-triggered price manipulation. (The presence of

standard price manipulation can be ruled out since  $G$  is positive semidefinite as the Fourier transform of  $\mu = \frac{1}{2}(\delta_{-\rho} + \delta_{\rho})$ . Using that

$$\int_0^{\pi/2} \cos(t-s) \sin(s) \, ds = \frac{1}{2} \cos(t) + \frac{\pi}{4} \cos\left(\frac{\pi}{2} - t\right),$$

we obtain

$$\begin{aligned} \int_{[0,T]} \cos(\rho(t-s)) \, dX_s^1 &= \int_0^{\pi/2} \cos(t-s) \left( -\frac{1}{2} \delta_0(ds) + \sin(s) ds - \frac{\pi}{4} \delta_{\pi/2}(ds) \right) \\ &= -\frac{1}{2} \cos(t) + \frac{1}{2} \cos(t) + \frac{\pi}{4} \cos\left(\frac{\pi}{2} - t\right) - \frac{\pi}{4} \cos\left(t - \frac{\pi}{2}\right) = 0, \end{aligned}$$

so  $X^1$  is a solution of (7.9). Putting

$$x_1 = - \int_0^{\pi/2} dX_s^1 = \frac{1}{2} - \int_0^{\pi/2} \sin(s) \, ds + \frac{\pi}{4} = \frac{\pi}{4} - \frac{1}{2},$$

$X^1$  is an optimal order execution strategy in  $\mathcal{X}(x_1, [0, \frac{\pi}{2}])$ . From the proof of Corollary 2, it follows that its time reversal

$$\tilde{X}_t^1 := \begin{cases} x_1 - X_{(\pi/2-t)+}^1 & \text{for } 0 \leq t \leq \frac{\pi}{2} \\ 0 & \text{for } t > \frac{\pi}{2} \end{cases}$$

is an optimal strategy in  $\mathcal{X}(x_1, [0, \frac{\pi}{2}])$  as well. But then also the symmetric strategy

$$\bar{X}_t^1 := \frac{1}{2} (X_t^1 + \tilde{X}_t^1)$$

has zero cost, meaning that we have found a third optimal execution strategy in  $\mathcal{X}(x_1, [0, \frac{\pi}{2}])$ .

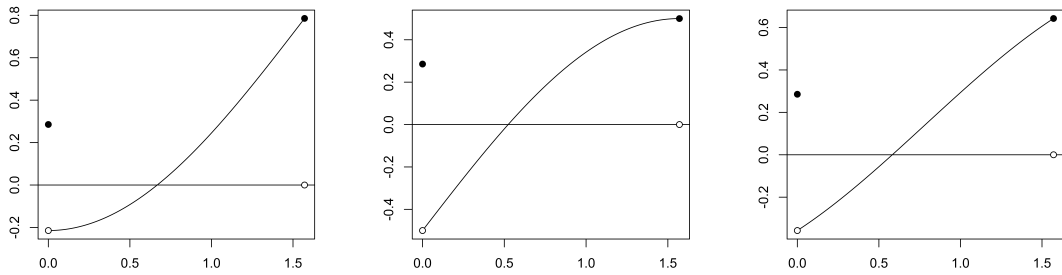


Figure 7.3: Optimal order execution strategies for the trigonometric decay kernel  $G(t) = \cos(t)$  on  $\mathbb{T} = [0, T]$  with  $X_0 = \frac{\pi}{4} - \frac{1}{2}$  and  $T = \frac{\pi}{2}$ . From left to right:  $X^1$ ,  $\tilde{X}^1$  and  $\bar{X}^1$ .

We can find even more optimal strategies by generalizing  $X^1$  to processes given by

$$dX_s^k = \frac{\sin(k\pi/2) - k}{k^2 - 1} \delta_0(ds) + \sin(ks) \, ds + \frac{k \cos(k\pi/2)}{k^2 - 1} \delta_{\pi/2}(ds)$$

for  $|k| \neq 1$  (cf. Figure 7.4). Using that

$$\int_0^{\pi/2} \cos(t-s) \sin(ks) \, ds = \frac{1}{k^2-1} \left( (k - \sin(k\pi/2)) \cos(t) - k \cos(k\pi/2) \cos\left(\frac{\pi}{2} - t\right) \right),$$

we obtain

$$\begin{aligned} & \int_{[0,T]} \cos(\rho(t-s)) \, dX_s^k \\ &= \int_0^{\pi/2} \cos(t-s) \left( \frac{\sin(k\pi/2) - k}{k^2-1} \delta_0(ds) + \sin(ks) \, ds + \frac{k \cos(k\pi/2)}{k^2-1} \delta_{\pi/2}(ds) \right) \\ &= \frac{\sin(k\pi/2) - k}{k^2-1} \cos(t) + \frac{k \cos(k\pi/2)}{k^2-1} \cos\left(t - \frac{\pi}{2}\right) \\ &\quad + \frac{1}{k^2-1} \left( (k - \sin(k\pi/2)) \cos(t) - k \cos(k\pi/2) \cos\left(\frac{\pi}{2} - t\right) \right) = 0, \end{aligned}$$

so  $X^k$  indeed solves (7.9). Putting<sup>5</sup>

$$\begin{aligned} x_k &:= - \int_0^{\pi/2} dX_s^k = \frac{k - \sin(k\pi/2)}{k^2-1} - \int_0^{\pi/2} \sin(ks) \, ds - \frac{k \cos(k\pi/2)}{k^2-1} \\ &= \frac{k - \sin(k\pi/2) - k \cos(k\pi/2)}{k^2-1} - \frac{\cos(k\pi/2) - 1}{k} \\ &= \frac{k^2 - k \sin(k\pi/2) - k^2 \cos(k\pi/2)}{k(k^2-1)} - \frac{(k^2-1) \cos(k\pi/2) - (k^2-1)}{k(k^2-1)} \\ &= \frac{1 - k \sin(k\pi/2) - \cos(k\pi/2)}{k(k^2-1)}, \end{aligned}$$

which is nonzero for almost all  $k$ ,  $X^k$  is an optimal execution strategy in  $\mathcal{X}(x_k, [0, \frac{\pi}{2}])$ . As for  $X^1$ , we can define the time reversal  $\tilde{X}^k$  of  $X^k$  and the symmetric strategy  $\bar{X}^k = \frac{1}{2}(X^k + \tilde{X}^k)$ , which are optimal as well. Since  $x_k \neq 0$  for almost every  $k$ , when dividing these strategies by  $x_k$ , we obtain a continuum of optimal order execution strategies in  $\mathcal{X}(1, [0, \frac{\pi}{2}])$ . In Figure 7.4,  $X^k/x_k$  is displayed for some values of  $k$ . Here it used that

$$\begin{aligned} X_t^k &= - \int_t^{\pi/2} dX_s^k = - \frac{k \cos(k\pi/2)}{k^2-1} - \int_t^{\pi/2} \sin(ks) \, ds \\ &= - \frac{k \cos(k\pi/2)}{k^2-1} + \frac{\cos(k\pi/2)}{k} - \frac{\cos(kt)}{k}. \end{aligned}$$

Note how the optimal strategy explodes for  $k = 4$ , which corresponds to  $x_4 = 0$ .  $\triangle$

---

<sup>5</sup>In [18], there is a misprint in the last equation.

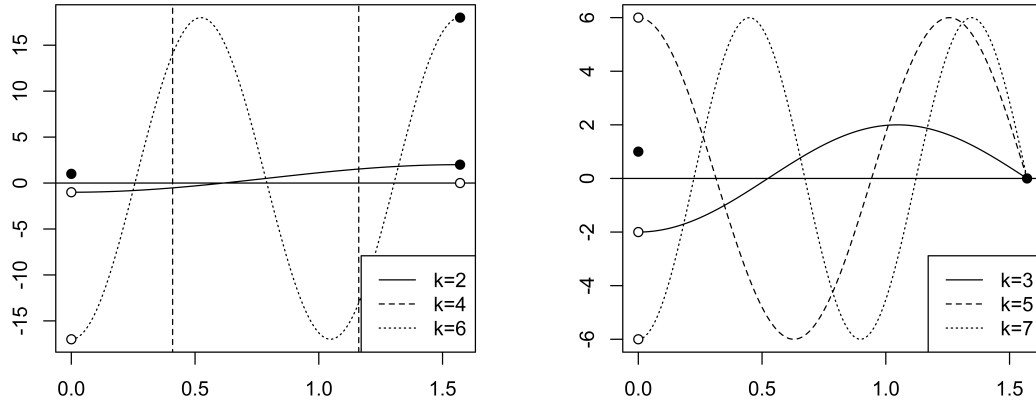


Figure 7.4: Optimal order execution strategies  $X^k/x_k$  with  $X_0/x_0 = 1$  for the trigonometric decay kernel  $G(t) = \cos(t)$  on  $\mathbb{T} = [0, T]$  with  $T = \frac{\pi}{2}$ .



## Chapter 8

# A well-behaved optimal trading strategy

In the previous chapter we found necessary conditions for the existence of a continuous-time optimal trading strategy. However, as opposed to the discrete-time case, we found that positive semidefiniteness of the decay kernel is not sufficient for an optimal strategy to exist. The question arises whether we can give conditions under which an optimal strategy does exist. In Chapter 4, we argued that optimal strategies should not alternate between buy and sell orders. In terms of a continuous-time strategy  $X$ , this means that  $X$  should be a monotone function of time. This gives rise to our second question: under which conditions is an optimal strategy a monotone function of time?

In this chapter, we will state conditions under which the existence of a unique optimal trading strategy is guaranteed. This optimal strategy will in fact appear to be a monotone function of time, thereby excluding transaction-triggered price manipulation as defined below. This is due to a compactness property of the class of monotone strategies in  $\mathcal{X}(x, \mathbb{T})$ , which can be used to prove the existence of optimal strategies (see the proof of Theorem 7 below).

As for discrete-time strategies, a market impact model is said to admit transaction-triggered price manipulation when the expected cost of a sell (buy) program can be decreased by intermediate buy (sell) trades (cf. Definition 2.19 in [18]). That is, there must be an admissible strategy  $\tilde{X}$ , supported on some compact set  $\mathbb{T}$ , such that<sup>1</sup>

$$\mathbb{E}[C(\tilde{X})] < \inf \left\{ \mathbb{E}[C(X)] \mid X \text{ admissible and monotone with support in } \mathbb{T} \text{ and } X_0 = \tilde{X}_0 \right\}.$$

We now state the theorem that guarantees the existence of an optimal strategy and the absence of transaction-triggered price manipulation, which is the main result in continuous-time modeling with nonsingular kernels. The proof of this theorem is based on Theorem 2, and the authors of [4] emphasize that they are “not aware of any argument that could bypass the discrete-time case” ([4]).

**Theorem 7.** ([18]) *Let  $G$  be a nonconstant nonincreasing convex decay kernel satisfying assumption (7.2). Then there exists a unique optimal strategy  $X^*$  within each class  $\mathcal{X}(x, \mathbb{T})$ . Moreover,  $X_t^*$  is a monotone function of  $t$ . That is, there is no transaction-triggered price manipulation.*

---

<sup>1</sup>In [18], the infimum is forgotten.

*Proof.* Since  $G$  is positive definite (as it is convex), by Theorem 4 the trivial strategy is the unique optimal strategy for the case  $x = 0$ . We can therefore assume that  $x \neq 0$ .

We have assumed that the set  $\mathbb{T}$  is compact, which means that it is a compact metric space with respect to the subspace topology induced by the embedding  $\mathbb{T} \subset \mathbb{R}$ . It follows that  $\mathbb{T}$  is separable, i.e. it admits a countable dense subset  $\{t_0, t_1, \dots\}$ . Using this, we define the finite set

$$\mathbb{T}_N := \{t_0, t_1, \dots, t_N\} \quad \text{for } N \in \mathbb{N}.$$

Now from Theorem 2, we know that for each  $N \in \mathbb{N}$  and given  $x$ , there exists a unique optimal strategy  $X^N$  with support on the subset  $\mathbb{T}_N \subset \mathbb{T}$ . Moreover, this optimal strategy  $X^N$  in the class  $\mathcal{X}(x, \mathbb{T}_N)$  is a nonincreasing or nondecreasing function of time, depending on the sign of  $x$ . Clearly,  $\mathcal{X}(x, \mathbb{T}_N)$  is a subset of  $\mathcal{X}(x, \mathbb{T})$ . Normalizing the strategy  $X^N$ , we thus get a Borel probability measure  $-\frac{1}{x}dX^N$  on  $\mathbb{T}$ . (Note that  $x$  and the jumps  $dX^N$  always have opposite sign.)

Now the family of Borel probability measures on  $\mathbb{T}$  is tight, i.e. for every  $\varepsilon > 0$  there exists a compact set  $K$  such that  $P(K) > 1 - \varepsilon$  for all  $P$  (simply by taking  $K = \mathbb{T}$ ). From Prohorov's theorem (cf. Theorem 6.1 in [10]), it follows that the space of all Borel probability measures on  $\mathbb{T}$  is relatively compact, meaning that every sequence  $(X^N)$  has a subsequence  $(X^{N_k})$  which converges to some (clearly monotone) strategy  $X^*$  in terms of weak convergence of the associated probability measures. Remark that this is where there the monotonicity of  $X^N$  plays a crucial role. Without it,  $-\frac{1}{x}dX^N$  becomes a signed measure, and we do not have relative compactness. We thus cannot guarantee the existence of the weak limit  $X^*$  when there is transaction-triggered price manipulation in the optimal execution problem on  $\mathbb{T}_N$ .

The next thing we show is that  $(X^{N_k})$  converges to  $X^*$  in terms of costs as well, that is, we prove that  $C(X^{N_k}) \rightarrow C(X^*)$  for  $k \rightarrow \infty$ . To this end, we observe that

$$C(X) = \int \int G(|t - s|) dX_s dX_t$$

is the integral of  $G(|t - s|)$  with respect to the product measure  $dX \otimes dX$ . Recall that  $G(s, t) := G(|t - s|)$  is bounded by assumption and continuous as a convex function of time. Now  $\mathbb{T} \times \mathbb{T}$  is separable as the product of separable spaces, and hence by Theorem 3.2 in [10] we get that

$$\int \int G(|t - s|) d(X_s^{N_k} \times X_t^{N_k}) \rightarrow \int \int G(|t - s|) d(X_s^* \times X_t^*),$$

or equivalently,

$$C(X^{N_k}) = \int \int G(|t - s|) dX_s^{N_k} dX_t^{N_k} \rightarrow \int \int G(|t - s|) dX_s^* dX_t^* = C(X^*).$$

It now remains to show that the limiting strategy  $X^*$  is optimal, i.e.  $C(X^*) \leq C(Y)$  for any strategy  $Y \in \mathcal{X}(x, \mathbb{T})$ . Such an arbitrary strategy  $Y$  can be characterized as  $Y = Y^+ - Y^-$ , where  $Y^\pm \in \mathcal{X}(x^\pm, \mathbb{T})$  are two nonincreasing strategies (i.e.  $x^\pm \geq 0$ ). Now since the set of Borel probability measures with support in  $\{t_0, t_1, \dots\}$  is dense in the set of all Borel probability measures on  $\mathbb{T}$ , there exist strategies  $Y^{\pm, N} \in \mathcal{X}(x^\pm, \mathbb{T}_N)$  which converge to  $Y^\pm$  in terms of weak convergence of the associated probability measures. Denoting

$$Y^N := Y^{+, N} - Y^{-, N} \in \mathcal{X}(x, \mathbb{T}_N),$$

and using (7.4) twice, we get that

$$\begin{aligned}
C(Y^N) &= C(Y^{+,N}) + C(-Y^{-,N}) + 2 C(Y^{+,N}, -Y^{-,N}) \\
&= C(Y^{+,N}) + C(Y^{-,N}) - 2 C(Y^{+,N}, Y^{-,N}) \\
&= 2 C(Y^{+,N}) + 2 C(Y^{-,N}) - C(Y^{+,N} + Y^{-,N}).
\end{aligned}$$

We have now expressed the cost of  $Y^N$  as a sum of costs of monotone strategies. By the same reasoning as above, these costs converge, yielding

$$C(Y^N) \rightarrow 2 C(Y^+) + 2 C(Y^-) - C(Y^+ + Y^-).$$

Using (7.4) two more times, we get back at

$$\begin{aligned}
2 C(Y^+) + 2 C(Y^-) - C(Y^+ + Y^-) &= C(Y^+) + C(Y^-) - 2 C(Y^+, Y^-) \\
&= C(Y^+ - Y^-) \\
&= C(Y),
\end{aligned}$$

so  $C(Y^N) \rightarrow C(Y)$ . Since  $X^N$  is optimal in  $\mathcal{X}(x, \mathbb{T}_N)$ ,  $C(X^N) \leq C(Y^N)$  for all  $N$ , so  $C(X^*) \leq C(Y)$  and  $X^*$  is an optimal strategy. Uniqueness follows by Theorem 4.  $\square$

From Theorem 7, we get that optimal strategies always exist and are well-behaved when price impact decays as a convex function of time. Conversely, when convexity is violated in a neighborhood of zero, it follows from Proposition 5 that there exist transaction-triggered price manipulation strategies.

**Proposition 14.** ([18]) *Suppose that there are  $s, t > 0$ ,  $s \neq t$  such that*

$$G(0) - G(s) < G(t) - G(t + s).$$

*Then there is transaction-triggered price manipulation for the choice  $\mathbb{T} = \{0, s, t + s\}$ .*

*Proof.* See the proof of Proposition 5.  $\square$

Similar to the discrete-time case, Theorem 7 and Proposition 14 together single out the class of nonnegative, nonincreasing, nonconstant, convex functions as the class of decay kernels for which the corresponding market impact model is sufficiently well-behaved. This brings us to the end of our study of transient linear price impact models in continuous time. We conclude this chapter with the continuous-time equivalent of Proposition 7.

**Proposition 15.** ([18]) *Let  $G$  be a nonconstant nonincreasing convex decay kernel satisfying assumption (7.2) and suppose that  $G'(0)$  is finite and  $x \neq 0$  (where  $G'$  is the right-hand derivative). Then the optimal strategy  $X^*$  in  $\mathcal{X}(x, \mathbb{T})$  has impulse trades at  $t_{\min} := \min \mathbb{T}$  and  $t_{\max} := \max \mathbb{T}$ , that is,  $\Delta X_{t_{\min}}^* \neq 0$  and  $\Delta X_{t_{\max}}^* \neq 0$ .*

*Proof.* See the proof of Theorem 2.23 in [18].  $\square$



# Conclusion

When the decay of price impact is given as a positive-definite quadratic form, the discrete-time second-generation market impact model introduced in [4] appeared to have a deterministic and unique optimal solution, which was given explicitly. Examples of the associated resilience functions showed a type of price manipulation different from standard price manipulation, which was called transaction-triggered price manipulation. To exclude this kind of price manipulation, convexity of the resilience function appeared to be both necessary and sufficient. Since nonconstant, convex functions generate positive definite quadratic forms, standard price manipulation was excluded in this case as well. The effects of risk aversion could be handled similarly to the way the standard optimal order execution problem was solved.

For the continuous-time version of the market impact model, it appeared that optimal strategies could be characterized as measure-valued solutions of a generalized Fredholm integral equation of the first kind. However, to guarantee the existence of an optimal trading strategy, positive definiteness did not hold, and we needed convexity of the decay kernel. As in the discrete-time case, this excluded the existence of transaction-triggered price manipulation strategies.



# Recommendations

This paper serves as an introduction to the concept of market impact modeling from a mathematical perspective. It covers the discrete-time transient linear price impact model introduced in [4], and its continuous-time equivalent introduced in [18].

Transient price impact models are also called second-generation models, to distinguish them from a first generation of models lacking a transient price impact component. In the introduction (Chapter 1) we had a very quick look at the first-generation model introduced in [6]. Similar models are treated in [5] and [7] and could be interesting to study. An overview of the theory on these first-generation models can be found in [30]. Examples of related concepts are maximization of expected utility, optimal control theory and Hamilton-Jacobi-Bellman equations.

Both in the discrete-time and continuous-time model that we studied, price impact was modeled linearly. Models with nonlinear price impact are studied in i.a. [1], [3] and [16]. For a discussion of the relation between linear and nonlinear models, one could consult [17]. In the recent paper [2], multivariate transient price impact is treated.

As mentioned in [4], the problem of excluding transaction-triggered price manipulation is closely related to the positive portfolio problem of characterizing the absence of short sales in a Markowitz portfolio. This latter problem is studied in i.a. [9], [20] and [24].

In Chapter 7 we mentioned that for practical implementation, it might be interesting to study the maximization of expected utility, leading to optimal strategies that are adapted to the stock price process. This is carried out in e.g. [28]. We also remarked that in [18], some of the results are extended so as to include weakly singular decay kernels. Being related to the concept of capacities in potential theory, the study of these weakly singular decay kernels is an exciting topic for further reading.

Market impact can also be approached from the empirical point of view. In Chapter 3, we mentioned that several studies (i.a. [12]) have found that price impact decays asymptotically with a power law. This topic is also covered in [16].





# Appendix A

## R code

In this appendix, we list the R code used to create the presented figures.

*Figure 2.1:*

```
xx=(300:1000)/100
xy=(150:300)/100
plot(xx,exp(-xx)+0.02,xlim=c(0,10),ylim=c(-0.12,0.12),type='l',xlab='',
      ylab='',lwd=3,xaxt='n',yaxt='n')
lines(xx,0*xx)
lines(xx,-exp(-xx)-0.02,lwd=3)
lines(xy,0*xy,lwd=3)
lines(xy,0*xy+0.11,lty=2)
lines(xy,0*xy+exp(-3)+0.02,lty=2)
lines(xy,0*xy+0.02,lty=2)
lines(xx,0*xx+0.02,lty=2)
lines(xy,0*xy-0.11,lty=2)
lines(xy,0*xy-exp(-3)-0.02,lty=2)
lines(xy,0*xy-0.02,lty=2)
lines(xx,0*xx-0.02,lty=2)
points(3,0,pch=19)
points(3,exp(-3)+0.02,pch=21,bg='white')
points(3,-exp(-3)-0.02,pch=21,bg='white')
points(3,0.11,pch=21,bg='white')
points(3,-0.11,pch=21,bg='white')
text(9,0.11,labels='BUY')
text(9,-0.11,labels='SELL')
arrows(8,0,10,0,length=0.1)
text(0.5,-0.11,labels='x G(0)')
text(0.5,0.11,labels='x G(0)')
text(0.5,-0.02-exp(-3),labels='x G(0+)')
text(0.5,0.02+exp(-3),labels='x G(0+)')
text(0.5,-0.02,labels=expression(paste("x G(", infinity, ")")))
text(0.5,0.02,labels=expression(paste("x G(", infinity, ")")))
abline(v=1.5)
```

#-----

```

nn=1000
nn1=2000
xx=(0:nn)/100
xx1=(nn:nn1)/100
yy=0*xx
yy1=0*xx1

for (k in 1:nn)
{
  yy[k+1] = yy[k] + rnorm(1,0,0.001)
  yy1[k+1] = yy1[k] + rnorm(1,0,0.001)
}

plot(xx,yy,"l",xlim=c(0,20),ylim=c(-0.1,0.1),xlab='',ylab='',
      xaxt='n',yaxt='n')
zz1=yy[nn+1]+yy1+exp(-(xx1-nn/100+3))+0.02
zz2=yy[nn+1]+yy1-exp(-(xx1-nn/100+3))-0.02
lines(xx1,zz1)
lines(xx1,zz2)
points(10,yy[nn],pch=19)
points(10,zz1[1],pch=21,bg='white')
points(10,zz2[1],pch=21,bg='white')
text(18,0.09,labels='BUY')
text(18,-0.09,labels='SELL')

```

*Figures 3.1, 3.3, 3.5, 3.7, 3.9, 3.11, 6.1, 6.2:*

```

X_0 = 10
N = 20
T = 10
sigma=0.3
gamma=5
e=exp(1)
rho= 0.5

tt = (0:N)/N*T
G = 0*(0:N)%*%t(0:N)
phi = G
one = rep(1,N+1)

for(i in 0:N)
{
  for(j in 0:N)
  {
    abst = abs(tt[i+1]-tt[j+1])
    #G[i+1,j+1]=exp(-abst) #exp.res
    #G[i+1,j+1]=e*(e-cos(abst))/(1+e^2-2*e*cos(abst)) #per.res
    #G[i+1,j+1]=(1+abst)^(-gamma) #PL
    #G[i+1,j+1]=max(1-rho*abst,0) #cap.lin.
  }
}

```

```

        #G[i+1,j+1]=1/(1+abst^2)                                #alt.PL
        #G[i+1,j+1]=exp(-abst^2)                                #Gaussian
        #G[i+1,j+1]=(1+abst)^(-0.4)                             #risk av.
        #phi[i+1,j+1]=sigma^2*min(c(tt[i+1],tt[j+1]))^(1/5)    #risk av.1
        #phi[i+1,j+1]=sigma^2*min(c(tt[i+1],tt[j+1]))          #risk av.2
    }
}

mbar = 1/2*G+gamma/2*phi
minv1 = solve(mbar)%*%one
xstar = X_0/((t(one)%*%minv1)[1,1])*minv1

barplot(t(xstar),ylim=c(0,8),names.arg=tt,axis.lty=1, xlab="N=20")
#xlab=expression(paste(rho,"=0.5">"1/T"))
abline(h=0)
box()

```

*Figures 3.2, 3.4, 3.8, 3.10:*

```

xx = (0:400)/100
e = exp(1)
#yy = e*(e-cos(xx))/(1+e^2-2*e*cos(xx))
#yy=(1+xx)^(-0.4)
#zz=(1+xx)^(-2)
#yy=1/(1+xx^2)
#yy=exp(-xx^2)

plot(xx,yy,"l",xlab="",ylab="",ylim=c(0,1))
#lines(xx,zz)

```

*Figure 3.6:*

```

xx1 = (0:500)/100
xx2 = (500:1000)/100
yy = (1-0.2*xx1)
plot(xx1,yy,"l",xlim=c(0,10),xlab="",ylab="",lwd=3,xaxt='n')
mtext(expression(paste("1/",rho)),1,0.5)
lines(xx2,0*xx2,lwd=3)
lines(xx1,0*xx1)

```

*Figure 4.1:*

```

xx = (0:1000)/100
yy=(6.5-atan(2*xx-12))/8-0.03*xx
plot(xx,yy,"l",xlab="",ylab="",xlim=c(-2,10),ylim=c(0.3,1.1),
     xaxt='n', yaxt='n')
points(0,(6.5-atan(2*0-12))/8-0.03*0,pch=19)

```

```

points(2,(6.5-atan(2*2-12))/8-0.03*2,pch=19)
points(5,(6.5-atan(2*5-12))/8-0.03*5,pch=19)
points(7,(6.5-atan(2*7-12))/8-0.03*7,pch=19)
xx1 = (0:200)/100
lines(xx1,((6.5-atan(2*2-12))/8-0.03*2)*(xx1/xx1),lty=2)
xx2 = (0:500)/100
lines(xx2,((6.5-atan(2*5-12))/8-0.03*5)*(xx2/xx2),lty=2)
xx3 = (0:700)/100
lines(xx3,((6.5-atan(2*7-12))/8-0.03*7)*(xx3/xx3),lty=2)
abline(v=0,h=0)
text(-1,1,labels="G(0)")
text(-1,0.935,labels="G(t)")
text(-1,0.8,labels="G(s)")
text(-1,0.465,labels="G(s+t)")

```

*Figure 7.1:*

```

xx = 10
TT = 20
rho = 0.1
tt=(0:200)/10
Xopt=xx*(rho*(TT-tt)+1)/(rho*TT+2)

plot(tt,Xopt,"l",ylim=c(0,10),xlab='',ylab='')
points(0,10,pch=19)
points(0,7.5,pch=21,bg='white')
points(20,2.5,pch=19)
points(20,0,pch=21,bg='white')

```

*Figure 7.2:*

```

xx = -10
TT = 5.2 #5
rho = 1
NN = floor(TT*rho)

tt=(0:(10*TT))/10
xstar=0*tt

for (jj in 0:(10*TT))
{
  xcomp=0
  for (ii in 0:NN)
  {
    xcomp=xcomp+(1-ii/(NN+1))*((jj/10==ii/rho)+(jj/10==(TT-ii/rho)))
  }
  xstar[jj+1]=xcomp
}

```

```

xstar[3] = xstar[51]
xstar[13] = xstar[41]
#I have to set these values manually since R is having trouble admitting
  that 0.2=0.2 and 1.2=1.2.

xstar=-xx/(2+NN)*xstar

barplot(t(xstar),ylim=c(0,2))
mtext("0  1  2  3  4  5 ",1,1)
box()

```

*Figure 7.3:*

```

tt=(0:round(100*pi/2))/100
Xopt1=pi/4-cos(tt)
Xopt2=cos(pi/2-tt)-1/2

plot(tt,Xopt1,"l",xlab='',ylab='')
#plot(tt,Xopt2,"l",xlab='',ylab='')
#plot(tt,(Xopt1+Xopt2)/2,"l",xlab='',ylab='')

abline(h=0)
points(0,pi/4-0.5,pch=19)
points(0,pi/4-1,pch=21,bg='white')
#points(0,-0.5,pch=21,bg='white')
#points(0,pi/8-0.75,pch=21,bg='white')
points(pi/2,pi/4,pch=19)
#points(pi/2,0.5,pch=19)
#points(pi/2,0.25+pi/8,pch=19)
points(pi/2,0,pch=21,bg='white')

```

*Figure 7.4:*

```

k=3
Xopt4a=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
  *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)
k=5
Xopt4b=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
  *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)
k=7
Xopt4c=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
  *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)

plot(tt,Xopt4c,"l",lty=3,xlab='',ylab='')
lines(tt,Xopt4a)
lines(tt,Xopt4b,lty=2)
legend("bottomright",c("k=3","k=5","k=7"),lty=1:3)

```

```

abline(h=0)
points(0,1,pch=19)
points(0,Xopt4a[1],pch=21,bg='white')
points(0,Xopt4b[1],pch=21,bg='white')
points(0,Xopt4c[1],pch=21,bg='white')
points(pi/2,0,pch=19)

#-----

k=2
Xopt4a=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
      *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)
k=4
Xopt4b=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
      *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)
k=6
Xopt4c=(-k*cos(k*pi/2)/(k^2-1)-cos(k*tt)/k+cos(k*pi/2)/k)
      *k*(1-k^2)/(cos(k*pi/2)+k*sin(k*pi/2)-1)

plot(tt,Xopt4c,"l",lty=3,xlab='',ylab='')
lines(tt,Xopt4a)
lines(tt,Xopt4b,lty=2)
legend("bottomright",c("k=2","k=4","k=6"),lty=1:3)

abline(h=0)
points(0,1,pch=19)
points(0,Xopt4a[1],pch=21,bg='white')
points(0,Xopt4b[1],pch=21,bg='white')
points(0,Xopt4c[1],pch=21,bg='white')
points(pi/2,Xopt4a[round(100*pi/2)+1],pch=19)
points(pi/2,Xopt4c[round(100*pi/2)+1],pch=19)
points(pi/2,0,pch=21,bg='white')

```

# Bibliography

- [1] Alfonsi, A., Fruth, A. & Schied, A. (2010), Optimal execution strategies in limit order books with general shape functions, *Quantitative Finance* **10**, pp. 143-157.
- [2] Alfonsi, A., Klöck, F. & Schied, A. (2014), Multivariate transient price impact and matrix-valued positive definite functions. Retrieved from <http://arxiv.org/pdf/1310.4471v4.pdf> (as of May 24, 2015)
- [3] Alfonsi, A. & Schied, A. (2010), Optimal trade execution and absence of price manipulations in limit order book models, *SIAM Journal on Financial Mathematics* **1**, pp. 490-522.
- [4] Alfonsi, A., Schied, A. & Slynko, A. (2012), Order book resilience, price manipulation and the positive portfolio problem, *SIAM Journal on Financial Mathematics* **3**, pp. 511-533.
- [5] Almgren, R. (2003), Optimal execution with nonlinear impact functions and trading-enhanced risk, *Applied Mathematical Finance* **10**, pp. 1-18.
- [6] Almgren, R. & Chriss, N. (1999), Value under liquidation, *Risk* **12**, pp. 61-63.
- [7] Almgren, R. & Chriss, N. (2000), Optimal execution of portfolio transactions, *Journal of Risk* **3**, pp. 5-39.
- [8] Barnett, S. (1990), *Matrices: methods and applications*, 1st edition, Clarendon Press, Oxford, pp. 361-364.
- [9] Best, M.J. & Grauer, R.R. (1992), Positively weighted minimum-variance portfolios and the structure of asset expected returns, *The Journal of Financial and Quantitative Analysis* **27**, pp. 513-537.
- [10] Billingsley, P. (1968), *Convergence of probability measures*, 1st edition, John Wiley & Sons, New York.
- [11] Breiman, L. (1992), *Probability*, 1st edition, Society for Industrial and Applied Mathematics, Philadelphia, p. 171.
- [12] Bouchaud, J.-P., Gefen, Y., Potters, M., & Wyart, M. (2004), Fluctuations and response in financial markets: the subtle nature of ‘random’ price changes, *Quantitative Finance* **4**, pp. 176-190.
- [13] Chang, K.-F. (1996), Strictly positive definite functions, *Journal of Approximation Theory* **87**, pp. 148-158.

- [14] CFTC-SEC (2010), Findings regarding the market events of May 6, 2010. Retrieved from <http://www.cftc.gov/ucm/groups/public/@otherif/documents/ifdocs/staff-findings050610.pdf> (as of April 16, 2015)
- [15] Folland, G.B. (1999), *Real analysis; modern techniques and their applications*, 2nd edition, John Wiley & Sons, New York.
- [16] Gatheral, J. (2010), No-dynamic-arbitrage and market impact, *Quantitative Finance* **10**, pp. 749-759.
- [17] Gatheral, J., Schied, A. & Slynko, A. (2011), Exponential resilience and decay of market impact. In *Econophysics of order-driven markets* (F. Abergel, B.K. Chakrabarti, A. Chakraborti, M. Mitra, eds.), Springer-Verlag, pp. 225-236.
- [18] Gatheral, J., Schied, A. & Slynko, A. (2012), Transient linear price impact and Fredholm integral equations, *Mathematical Finance* **22**, pp. 445-474.
- [19] Gnedenko, B.V. (1997), *Theory of probability*, 6th edition, Overseas Publishers Association, Amsterdam.
- [20] Green, R.C. (1986), Positively weighted portfolios on the minimum-variance frontier, *Journal of Finance* **41**, pp. 1051-1068.
- [21] Karatzas, I. & Shreve S.E. (1998), *Brownian motion and stochastic calculus*, 2nd edition, Springer-Verlag, New York, p. 10.
- [22] Marsden, J.E. & Tromba, A.J. (2003), *Vector calculus*, 5th edition, W.H. Freeman and Company, New York, p. 226.
- [23] Meyer, C.D. (2000), *Matrix analysis and applied linear algebra*, 1st edidtion, Society for Industrial and Applied Mathematics, Philadelphia, p. 559.
- [24] Nielsen, L.T. (1987), Positively weighted frontier portfolios: a note, *Journal of Finance* **42**, p. 471.
- [25] Obizhaeva, A. & Wang, J. (2006), Optimal trading strategy and supply/demand dynamics. Retrieved from [http://web.mit.edu/wangj/www/pap/OW\\_060408.pdf](http://web.mit.edu/wangj/www/pap/OW_060408.pdf) (as of May 8, 2015).
- [26] Roberts, A.W. & Varberg, D.E. (1973), *Convex functions*, 1st edition, Academic Press, New York.
- [27] Rynne, B.P. & Youngson, M.A. (2008), *Linear functional analysis*, 2nd edition, Springer, London, p. 17.
- [28] Schied, A. & Schöneborn, T. (2009), Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets, *Finance and Stochastics* **13**, pp. 181-204.
- [29] Schied, A., Schöneborn, T. & Tehranci, M. (2010), Optimal basket liquidation for CARA investors is deterministic, *Applied Mathematical Finance* **17**, pp. 471-489.
- [30] Schied, A. & Slynko, A. (2011), Some mathematical aspects of market impact modeling, *Surveys in Stochastic Processes. Proceedings of the 33rd SPA, EMS Series of Congress Reports*, pp. 153-179.



- [31] Stade, E. (2005), *Fourier analysis*, 1st edition, John Wiley & Sons, Hoboken, New Jersey.
- [32] Van Vliet, L.J., Rieger, B. & Verbeek, P.W., Fourier transform of a Gaussian. Retrieved from <http://homepage.tudelft.nl/e3q6n/education/tn254/2002/Fourier%20transform%20of%20a%20Gaussian.pdf> (as of March 26, 2015)
- [33] Weisstein, E.W., Fourier Transform–Exponential Function. From *MathWorld*—A Wolfram Web Resource. <http://mathworld.wolfram.com/FourierTransformExponentialFunction.html> (as of March 4, 2015)